DOCUMENT RESUME

ED 230 205 IR 050 272

AUTHOR	Russell, Keith W., Comp.
TITLE	Subject Access: Report of a Meeting Sponsored by the Council on Library Resources (Dublin, Ohio, June 7-9, 1982).
INSTITUTION	Council on Library Resources, Inc., Washington, D.C.
PUB DATE	Jun 82
NOTE	97p.
PUB TYPE	Viewpoints (120) -- Reports - Descriptive (141) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Cataloging; *Classification; Databases; Definitions; Indexing; *Information Retrieval; *Online Systems; *Subject Index Terms
IDENTIFIERS	*Bibliographic Data Bases; *Library of Congress Subject Headings

ABSTRACT

A meeting was held to explore critical issues and problems related to subject access to online bibliographic records of monographic literature. The 23 researchers, practitioners, and policymakers who attended represented national and research libraries, bibliographic utilities, database vendors, library schools, indexing and abstracting services, and other organizations. Recommendations for study and action in several areas (subject heading, classification, database, and general issues) were prepared by small working groups and refined by the entire group into a set of 16 recommendations to improve subject access for users. Priorities were also established to divide the recommendations into short-term, long-term, and other projects. This final meeting report includes five papers: (1) "Subject Access/Subject Authority," by C. Lee Jones; (2) "Subject Access in Library of Congress Catalog Records," by Lucia J. Rather and Mary K. Pietris; (3) "Affordable Enhancements to Bibliographic Records for Subject Access," by William Mischo; (4) "Word, Phrase and Term (Descriptor) Searching," by Elaine Svenonius; and (5) "Classification as an Online Subject Access Tool," by Pauline Cochrane. The 16 recommendations, which deal mainly with Library of Congress Subject Headings (LCSH), are also presented, as are working definitions of "subject access" suggested by the participants. The report further describes meeting agenda and discussions and reviews progress made relative to the recommendations since the meeting. (ESR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED230205

SUBJECT ACCESS

Report of a meeting sponsored by

The Council on Library Resources

Dublin, Ohio

June 7-9, 1982

Compiled and edited by Keith W. Russell

Council on Library Resources, Inc.
1785 Massachusetts Avenue, N.W.
Washington, D.C. 20036

December 1982

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Jane Rosenberg

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR050272

SUMMARY

A meeting sponsored by the Council on Library Resources to explore critical issues and problems related to subject access in bibliographic records of monographic literature was held in Dublin, Ohio, June 7-9, 1982. Twenty-three researchers, practitioners, and policy makers interested in the topic of subject access were brought together for this meeting; they represented national and research libraries, bibliographic service agencies (utilities), database vendors, library schools, index and abstract services, and other organizations.

The purpose of the meeting was to identify a set of long-term and short-term recommendations that could be acted upon by a variety of individuals and organizations to improve subject access in bibliographic databases for the user. Participants received background papers before the meeting, and four papers were given at the meeting. Recommendations for study and action in several areas (subject heading issues, classification issues, database issues, and general issues) were prepared by small working groups; these were then carefully refined by the entire group into a set of 16 recommendations--the end product of the meeting. Priorities were established for these recommendations, and they were divided into short-term projects (those that could be started and substantially completed within three years), long-term projects (those that could be started in the near term, but would require effort beyond three years), and other projects (deemed important, but for which priorities were not set).

This report presents those 16 specific recommendations, 6 assumptions that were made in arriving at those recommendations, one of the background papers, the discussion papers, and working definitions of "subject access" suggested by the participants. It describes how the meeting was conducted, who the participants were, and the discussions that took place. It also reports on progress that has been made relative to some of the recommendations since the meeting was held.

CONTENTS

Summary iii

Preface, by Warren J. Haas. ix

Chapter 1: Introduction. 1

Chapter 2: Background. 5

2.1 Subject Access / Subject Authority, by C. Lee Jones 6

Chapter 3: Getting Started 9

3.1 Setting the Stage 9
3.2 Subject Access in Library of Congress Catalog Records, by Lucia J. Rather and Mary K. Pietris. 10
3.3 Affordable Enhancements to Bibliographic Records for Subject Access, by William Mischo 19
3.4 Word, Phrase and Term (Descriptor) Searching, by Elaine Svenonius 30
3.5 Classification as an Online Subject Access Tool, by Pauline Cochrane 35

Chapter 4: Developing Recommendations. 41

4.1 The Subgroups 41
4.2 Recommendations of Subgroup A 43
4.3 Recommendations of Subgroup B 45
4.4 Recommendations of Subgroup C 47
4.5 Recommendations of Subgroup D 49

Chapter 5: Refining Recommendations and Setting Priorities 53

5.1 Assumptions 54
5.2 Subject Heading Issues. 55
5.3 Classification Issues 59
5.4 Database Issues 60
5.5 General Issues. 64

Chapter 6: Summary of Recommendations. 67

6.1 Assumptions 68
6.2 Short-term Projects/Issues. 69
6.3 Long-term Projects/Issues 71
6.4 Other Projects/Issues 72

Chapter 7: Epilogue. 75

7.1 The Wednesday Afternoon Meeting 75
7.2 Fostering Communication 75
7.3 LCSH in Machine-readable Form 76
7.4 Contributions to LCSH by Other Libraries. 76
7.5 LCSH Hierarchy. 77
7.6 Enhancing Subject Access. 79
7.7 Reactions to Recommendations. 79

Appendices

Appendix A: Agenda A-1

Appendix B: Participants B-1

Appendix C: Definitions of "Subject Access". C-1

Appendix D: The Wednesday Afternoon Meeting. D-1

PREFACE

This report of a meeting on subject access is both a product of a process and a careful record of the process itself. Readers thus have a context in which to judge the validity of recommendations and an opportunity to share vicariously in the conversations themselves. Lengthy minutes often stem from unfocused discussions but, in this case, the length of the record reflects intense and productive sessions that are fully reported because they deal with a matter of central importance--the usefulness of bibliographic systems and services to individuals, whether they be experienced scholars, students, beginning researchers, or any other member of the American public. As the volume of recorded information grows in quantity, becoming along the way more varied in format and infinitely more complex in content, it is imperative that better ways be found to enable would-be users to identify the specific items they need with reasonable ease and precision. Improved access to bibliographic records by subject is, for users, a matter of growing importance because of the sheer quantity of recorded information with which they must deal.

Computer technology has, fortunately, flourished in phase with the growing information store (and is itself, a reason for much of that growth). The work that is ahead concerns using technology to help resolve the difficult intellectual problem of linking the information requirements of individuals working today to the accumulated record of our intellectual past. To do so in a reliable, satisfactory, and economically feasible way is the task ahead. The discussions reported here mark an important step along the road.

Warren J. Haas

ACKNOWLEDGEMENTS

Appreciation is due the participants in the Subject Access Meeting, CLR staff, and all other interested persons who helped create this report.

CHAPTER 1

INTRODUCTION

For some time the Council on Library Resources has been interested in the topic of subject access to bibliographic records, and ways of improving that access for the end user. The best approaches to improvement have not been clear, but developments early in 1982 suggested it was time to re-focus attention on the problem. To help identify fruitful areas for study and action, CLR sponsored a meeting of knowledgeable persons in the area of subject access.

The meeting, an initiative of the Council's Bibliographic Service Development Program (BSDP), was held in Dublin, Ohio, beginning Monday evening, June 7, 1982, and ending at noon Wednesday, June 9. Daytime sessions were held at OCLC, and evening sessions at the nearby Stouffer's Dublin Hotel. Appendix A is the agenda of the meeting.

Twenty-three people--researchers, policy makers and practitioners--from national and research libraries, library schools, database vendors, indexing and abstracting services, bibliographic service agencies (utilities), and other organizations participated in the meeting. They were chosen from a long list of people known to be interested in subject access from a number of perspectives. Appendix B identifies the participants.

Three papers providing background information were distributed to the participants in advance. Citations to the two previously-published of these (by Carol Mandel and Judith Herschman, and William Mischo) are included in chapter 2, along with the text of the third, prepared by C. Lee Jones

specifically for the meeting. These background papers review research on subject access that could have implications for subject access in online catalogs, propose a sophisticated subject retrieval function for an online union catalog, and discuss some of the issues involved.

The first program session (Monday evening, June 7) was devoted to introducing the participants, reviewing past CLR efforts in the area of subject access, defining the problem, describing the purpose, mechanics and focus of the meeting, and discussing assumptions upon which the deliberations would be based. Definitions of "subject access" by participants (which they had been asked to bring to the meeting) were distributed and briefly discussed; they are reproduced in Appendix C. The Monday evening session is more fully described in Chapter 3.

On Tuesday morning four papers were presented by five of the participants to help build a common understanding and stimulate discussion. Topics covered included the current status of subject access in LC bibliographic records, ways of enhancing bibliographic records for subject access, and the use of sophisticated search techniques and classification schedules to improve subject access. Chapter 3 includes the text of these presentations and notes of discussions during the presentations.

Tuesday afternoon the participants worked in small groups to identify major issues and possible solutions, and to arrive at recommendations for actions that could help improve subject access for the user. Tuesday evening each subgroup reported on its deliberations and recommendations. Chapter 4 describes the process by which the recommendations were developed, and summarizes the reports of the subgroups.

Wednesday morning the recommendations from the subgroups were reviewed and refined, and priorities for action were set. Chapter 5 is a detailed

description of the process by which a final set of 16 recommendations and 6 assumptions were derived from the 40 recommendations and 7 assumptions contributed by the subgroups.

(Since this report is designed to be useful to a variety of readers, much detail about the process and the deliberations is included in chapters 4 and 5. Some readers may want to skip these chapters and focus more on the summary of recommendations in Chapter 6 and other parts of the report relevant to their interests.)

Chapter 6 is a summary of the final set of recommendations and assumptions resulting from the meeting, and is the most important chapter. Chapter 7 reports on developments related to those recommendations that have taken place in the six months since the Subject Access Meeting was held. Appendix D is a report of one of those developments: a meeting of some Subject Access Meeting participants with LC processing staff immediately after the main meeting.

* * * * * * *

CHAPTER 2

BACKGROUND

Three background papers were distributed in advance to meeting participants. Two of these had previously been published, and are not reproduced here. Those two items, available from the sponsoring agencies, are:

Mandel, Carol A., with the assistance of Judith Herschman. Subject access in the online catalog: a report prepared for the Council on Library Resources. Washington, D.C., Council on Library Resources, August 1981.

Mischo, William H. Technical report on a subject retrieval function for the online union catalog. Dublin, Ohio, OCLC Development Division, November 20, 1981. (Report number OCLC/DD/TR-81/4)

The third background paper, prepared by C. Lee Jones specifically for this meeting, is reproduced below.

2.1 BACKGROUND PAPER

SUBJECT ACCESS / SUBJECT AUTHORITY
Challenge and Opportunity

by C. Lee Jones

The Council's Bibliographic Service Development Program (BSDP) staff and advisors have identified the need for progress in the area of subject authorities and subject access for several years. However, the specific actions that need to be taken have not been identified, nor has the problem been sufficiently defined. On several occasions the Program Committee of the BSDP invited consultants to talk with them about subject access in an effort to define what is required, but results of these discussions were less than satisfactory.

Not until Carol Mandel with Judith Herschman produced their report Subject Access in the Online Catalog, did the BSDP have a clearly defined set of options for action in the area of subject access and related problems. Shortly after this document was presented, the preliminary results of the pilot study of online public access catalogs were reported at a symposium preceeding the ALA mid-winter meeting in Denver, January 1982. These results clearly indicated a concern on the part of users for enhanced and more simple means of accessing online catalogs via subjects or word terms.

Subject access has at least two general aspects, related but separate. Subject authority or subject control requires, or at least implies, a controlled vocabulary approach to the content of a bibliographic record. Some subject systems are based on thesauri which organize subject terms in a hierarchical fashion. Other systems, including the Library of Congress

Subject Headings (LCSH), are aggregations of subject terms not necessarily rigorously related to each other as in hierarchical systems.

A very separate topic is that of subject access--independent of the structure of the subject vocabulary used to create a record, but intimately involved with the methods available for searching the record. Users often define any search of an online public access catalog that is not specifically by a known title or author as a subject search. Therefore, for purposes of this discussion, any such search should also be referred to as a "subject" search. It is important to remember the context in which users use and view this bibliographic tool. The fact that the search is a word or related word search does not impress most users as anything but a subject search.

System designers must seriously consider abandoning the concept of "subject searches" as searches only of those fields containing authorized subject terms. The more useful approach would seem to be to provide the capacity to search all fields of a record for the terms input at the terminal by the user. This does not suggest that subject headings should no longer be assigned. On the contrary, it underscores the need for as useful a subject approach as can be devised. The core concept here is to combine the advantages of the structured list of subject headings, the controlled thesaurus, with the advantages of word or term searches. Users should not have to distinguish between the two. In fact, it is not clear why there should be a distinction for purposes of the usual subject approach to the online public access catalog. The argument that there should be a search capacity limited to whatever thesaurus structure is used in the record does not preclude a combined subject/term/word search capability for the less sophisticated user.

To date, few online public access catalogs have taken advantage of one of the more obvious subject approaches to online databases that are also organized by a classification structure that is subject based. Many sources, including ALA's committee concerned with subject access, have urged that this

approach be explored, and where feasible, implemented. The present study of online public access catalogs does include catalogs that have this feature.

Finally, there are a number of high quality thesauri used to provide controlled subject access to bibliographic records, including LC's LCSH, NLM's MeSH, and several smaller ones. The Art and Architecture librarians have just completed a survey of the several thesauri available in their specific area. It became clear that most of them are based on LC's work or that in order to be widely accepted, any alternative must relate to LC's work in some effective way. How can this be assured? It has been suggested that the Art and Architecture group that is developing a hierarchical thesaurus patterned after NLM's MeSH should work with LC in order to understand whether or not that portion of the LCSH might be amenable to conversion to a hierarchical structure. If such a conversion were possible in this area of LCSH, would a similar strategy be a viable possibility for the entire LCSH? What would such a conversion cost? Not only at LC, but at all those institutions that use LC's records? LC, though it may be willing, must still deal with the flow of material coming through on a daily basis, and would have to continue to do so during any conversion period.

What are the basic questions that need to be addressed in order to enhance subject access both controlled and free? How can the BSDP make an effective contribution to improving the subject access capabilities available to the users of online catalogs? The results of the online public access catalog study appear to be adding weight to the notion that something must be done in order to meet the subject access needs of catalog users.

* * * * * * *

CHAPTER 3

GETTING STARTED

This chapter reports on the opening session, Monday evening, June 7, which set the stage for the meeting, and the Tuesday morning session at which four papers were presented to provide a common foundation and to stimulate further discussion. The papers are reproduced in sections 3.2-3.5, along with a summary of discussions that took place. It is important to note that these papers were prepared over a short period of time, and were designed for informal presentation.

3.1 SETTING THE STAGE

The Monday evening session included introductions of participants (listed in Appendix B), a review of the Council's Bibliographic Service Development Program efforts in the area of subject access, description of the purpose and mechanics of the meeting (the agenda is given in Appendix A), and discussion of scope and assumptions.

The focus of the meeting was defined as subject access for the lay end user (and not the intermediary trained searcher). Issues to be considered would be limited to those that have some impact on how a user gets access to information about monographs in a bibliographic database. Wherever appropriate, anything bearing on access to both monographic and journal literature would be discussed, but technologies with applications only to the journal literature would be excluded since discussion time would be limited.

Discussion of assumptions underlying the meeting resulted in this listing: (a) discussions would relate only to searches by the end user; (b) the databases involved would consist of MARC records; (c) a controlled vocabulary such as the Library of Congress Subject Headings (LCSH) would be used; and (d) LCSH would not be eliminated. Discussions would focus on subject access to the 15 million records now available on the major systems, as well as subject access to records not yet created.

The definitions of "subject access" prepared by the participants were briefly discussed to help set the stage for further discussions. Those definitions are presented in Appendix C.

3.2 DISCUSSION PAPER 1

SUBJECT ACCESS IN LIBRARY OF CONGRESS CATALOG RECORDS

by Lucia J. Rather and Mary K. Pietris

The bibliographic record was designed 80 years ago to be used in a card catalog. The record includes a wealth of information, and it costs approximately $100 to create one at the Library of Congress (LC).

Subject headings are the traditional subject access points to bibliographic records. At the Library of Congress subject headings have usually been assigned to reflect the topic of a work as a whole, not to provide in-depth analysis. Subject headings are selected from the two-volume Library of Congress Subject Headings (LCSH), now in its ninth edition with three supplements. Subject headings are created daily as needed for new cataloging at the Library of Congress. LC now has systems which provide

keyword search capabilities within subject headings, greatly expanding the number of access points in each bibliographic record.

LC plans to develop a system which will permit input, update, and searching online. In a later stage, this system will permit distribution of a master database tape and regular update tapes. At present, it takes about one year for librarians in the field to get the paper supplements. In about a month a master database through the end of 1980 will be available on tape. However, the distribution of update tapes will not be possible for two to three years.

Each bibliographic record has an average of 1.996 subject headings (2.138 if you exclude class P (literature)), as seen below:

AACR 2 BOOKS STATISTICS
(Portion of 1982 titles cataloged)

Total file:	84,044	Total excluding class P:	69,202
Total number of 6XX	167,778		147,982
Avg hdgs per record:	1.996		2.138

In a card catalog one can only enter via the first word in a subject heading string. Records normally have only headings from LCSH, but medical books may also have Medical Subject Headings (MeSH), and childrens' books may have LC Childrens' Subject Headings.

The Library of Congress call number, consisting of the LC classification number and a book number, is another access point that could be used. The classification enables works on the same subject to be grouped together on

library shelves; the book number arranges works within a subject by main entry. Works classed in Z (bibliography) and works grouped together by series name have an additional class number assigned to reflect the subject.

Medical books may have both LC and NLM class numbers. Approximately 100,000 books cataloged by LC each year also have Dewey Decimal Classification numbers, primarily because Dewey is heavily used outside the U.S. and in smaller U.S. libraries. (The Dewey classification is hierarchical, whereas the LC classification is not; consequently Dewey has special features useful for subject access.) Another approach to subject access using the current bibliographic record is use of the geographic area code, which is assigned to reflect the geographic area covered by the subject of the book.

The Dewey Decimal Classification reflects the notation according to the system devised by Melville Dewey. Many works in English and in Western European languages are assigned Decimal numbers. Alternate numbers are assigned for works classed together under the series.

Unstructured approaches to bibliographic records are possible in a machine environment, where words embedded anywhere in a traditional bibliographic record can be searched:

1. Title and key words in title. The access in a card catalog reflects only the first word in the title and the partial title added entry, if any was assigned by the cataloger. With computers and with keyword search capabilities, any significant word in the title represents a subject access point, insofar as those words describe what the work is about and are searchable fields in the system.

2. Series entries, and key words within series. A series may also contain significant subject access points accessible by keyword searching.

3. Main and added entries, and key words within. Key words from main and added entries may also provide subject access. Automation has brought a new method of subject access by providing the ability to find words embedded in any part of the traditional bibliographic record if these words are indexed or free text searching is used.

4. Contents, notes and annotations, which are present in only 2-5% of records, can also provide additional subject access points.

The following bibliographic records (Figures 3.1 to 3.6) illustrate that while a record may have one, two, three or more subject headings assigned to it, those subject headings do not represent the only subject access points contained in records. Words underlined in each record are additional terms with subject meaning that are not searchable in the card catalog but can be searched in some online catalogs. The ability to search on those terms would greatly add to the subject access available to existing records.

FIGURE 3.1

CRD 81-104029
CAL Z7164.04S49 (BQ1063)
DDC 016.3052/6 19
MEP Sigel, Lois.
TIL Multigenerational considerations in planning environments for the elderly: an annotated bibliography / Lois Sigel.
IMP Monticello, Ill. : Vance Bibliographies, 1980.
COL 19 p. ; 28 cm.
SET Architecture series--bibliography , 0194-1356 ; A-395
NOG Cover title.
SUT Retirement communities Bibliography.
SBN $3.00 (pbk.)

FIGURE 3.2

CRD 81-81696
CAL D802.Y8D33 1981
DDC 940 53/45 19
MEP Davidson, Basil.
TIL Scenes from the anti-Nazi war / by Basil Davidson.
IMP New York : Monthly Review Press, (1981) c1980.
COL 228 p. : map ; 21 cm.
NOG Includes index.
NOB Bibliography: p. (279)-284.
SUP Davidson, Basil.
SUC Great Britain. Special Operations Executive Biography.
SUT World War, 1939-1945-Underground movements-Yugoslavia-Biography.
SUT World War, 1939-1945-Underground movements-Italy-Biography.
SUT World War, 1939-1945-Personal narratives, English.
SUT Guerrillas-Yugoslavia-Biography.
SUT Guerrillas-Italy-Biography.
SBN ISEN 0-85345-587-2
SBN ISBN 0-85345-588-0 (pbk.)
GAC e-uk--- e-yu--- e-it---

FIGURE 3.3

CRD 81-607149
CAL RA407.3.A349 no. 63 (RG106.4.U6)
DDC 362.1/1/0973 s 352.1/981/00973 19
MEP Cypress, Beulah K.
TIL Use of health services for disorders of the female reproductive system, United States, 1977-78.
IMP Hyattsville, Md. : U.S. Dept. of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health, National Center for Health Statistics, (1981)
COL p. cm.
SET Vital and health statistics. Series 13, Data from the national health survey ; no. 63
SET DHHS publication ; no. (PHS) 82-1724
NOB Includes bibliographical references.
SUT Generative organs, Female Diseases-United States-Statistics.
SUT Women's health services-United States-Utilization-Statistics.
SUT Physician services utilization-United States-Statistics.
SUT Gynecology, Operative-United States-Statistics.
SUT Breast Surgery-United States-Statistics.
SUT Health surveys-United States.
SUT Genital diseases, Female Therapy. (MeSH)
SUT Health services-Utilization-United States. (MeSH)
SUG United States-Statistics, Medical.
AEC National Center for Health Statistics (U.S.)
SDC HE 20.6209:13/63
SBN ISBN 0-8406-0244-8
GAC n-us---
NLM W2 A NI48vm no. 63

FIGURE 3.4

CRD 81-19854
CAL RC388.5.D64
DDC 616.8/107543 19
TIL Doppler ultrasound in the diagnosis of cerebrovascular disease / edited by Robert S. Reneman and Arnold P.G. Hoeks.
IMP New York : Wiley, c1982.
COL p. cm.
SET Ultrasound in biomedicine research series ; v. 5
NOG Includes index.
SUT Cerebrovascular disease- Diagnosis.
SUT Diagnosis, Ultrasonic.
SUT Doppler effect.
AEP Reneman, Robert S.
AEP Hoeks, Arnold P. G.
SBN ISBN 0-471-10165-6 : $59.00

FIGURE 3.5

CRD 81-16748
CAL BJ1421.B69
DDC 179/.8 19
MEP Bowyer, J. Barton.
TIL Cheating : deception in war & magic, games & sport, sex & religion, business & con games, politics & espionage, art & science / by J. Barton Bowyer.
IMP New York, N.Y. : St Martin's Press, 1982.
COL p. cm.
SUT Deception.
SBN ISBN 0-312-13167-4 : $11.95
SBN ISBN (invalid) 0-312-19008-5

FIGURE 3.6

CRD 81-19981
CAL TX724.5.B3067
DDC 641.5 19
MEF Corum, Ann Kondo.
TIT Easy cooking--the island way : a cookbook for singles, senior citizens, college students, newlyweds, teenagers, working people, and others who want nutritious, economical and uncomplicated good food / by Ann Kondo Corum.
IMP Kailua, HI : Press Pacifica, 1981.
COL p. cm.
NOG Includes index.
SUT Cookery, Hawaiian.
SBN ISBN 0-916630-24-2
GAC n-us-hi

* * * * * * *

Discussion

During discussion of the examples, the limitations of online systems that only allow searching the main title (not the subtitle) for keywords were mentioned. NLM has had experience providing additional subject access points by using class numbers in machine-readable form, along with brief phrases describing what each class number means. The name of this file is MEDCLASS, and evidence so far is that there is much potential for use of such files in online public access catalogs. The work Nancy Olsen has done at Mankato State using a word processor to input all indexes to the LC Classification System, thus providing links to class numbers, was also mentioned. Such applications can be very useful for increasing subject access to existing records; adding phrases that describe DDC and LCC numbers make such techniques even more useful.

Arthur D. Little, Inc., worked with Forest Press to produce a print tape of the 19th edition of the Dewey Decimal Classification. Forest Press is now exploring ways the machine-readable format could be used in other ways, keeping in mind that the DDC must be self-supporting. The R. R. Bowker Company has correlated LCC and DDC numbers and entered the equivalent subject headings that go with them; Ed O'Neill has also done work in this area.

One problem could be how to help users narrow the search if several additional points of access are provided. Legislative files often have a large amount of text--in many cases, 10 to 15 screens--and frequently a large number of false retrieved items occur when the full text of such files is searched for retrieval purposes. Problems involving such rich records are being solved, however, and the existence of these problems in no way reduces the need for, and value of, enriching poor records.

Searches that provide too many retrieved items can be narrowed in several ways, including use of: date, language, type of publication; "proximity operators"; major versus minor subject headings; Boolean operators; and class codes. Limiting the search is one of the top five problems in many online public access catalogs; one of the top two problems, however, is how to broaden a search that results in too few hits. Thirty-three percent of searches on MELVYL, the University of California, Division of Library Automation's online catalog, result in no hits, while twelve percent of searches result in over 100 hits. That online catalog contains about 750,000 records, and full-text access to the entire bibliographic record is provided.

3.3 DISCUSSION PAPER 2

AFFORDABLE ENHANCEMENTS TO BIBLIOGRAPHIC RECORDS FOR SUBJECT ACCESS

by William Mischo

I. Introduction

The inadequacies of Library of Congress Subject Headings (LCSH) in providing subject access to library materials has been well documented (1,2,3,4). Mischo's paper in Cataloging and Classification Quarterly identifies, by anecdote and case study, the reasons for subject retrieval failure.

Yet, catalog use studies (e.g., Lipetz) and recent surveys of users of online public access catalogs (OPACs) clearly show the importance of subject access and suggest that improved subject access is needed (5). This overview paper will focus on three areas offering the potential for enhanced subject access: 1) the use of online search techniques for accessing information

contained in catalog records; 2) augmentation of LCSH with additional subject descriptors/headings; and 3) improved subject authority services.

II. Online public access catalogs

In recent years, the capability of searching A&I (abstracting and indexing) service databases in an online mode has refined the art of information retrieval. Subject access to library materials can be improved, even without LCSH augmentation, by the capability of performing online searches of MARC record data. Improved subject access has been demonstrated by online systems using Boolean search operations and word truncation over title words, subject heading words, series title words, contents fields, and classification code fields. The importance of title word access in subject retrieval, as illustrated in figures 3.7 - 3.9, has been well documented (6,7,8). Figure 3.10 illustrates the value of access to words appearing in the notes field. Subject retrieval in the online catalog can be further enhanced by allowing limitation of search results by year of publication, language of publication, and type of material and qualification of search arguments to specific fields or sets of fields within the record. The online catalog, with its potential for sophisticated user/system dialogs, offers the opportunity for improved subject access through the use of the computer-guided search techniques (such as conducting subject searches in the title and subject heading fields) and suggestive prompts based on search results (such as displaying subject authority headings or classification schedules in response to searches retrieving a high number of postings). Figure 3.11 illustrates the value of searching multiple fields--in this case, the title and subject fields.

FIGURE 3.7

LO DRBU.
TL875-L963-1976.
ME Lyndon B. Johnson Space Center.
AE United States. National Aeronautics and Space Administration.
TI Space shuttle / prepared by Lyndon B. Johnson Space Center.
IM Washington: U. S. National Aeronautics and Space Administration, Scientific and Technical Information Office: for sale by the Supt. of Docs. U.S. Govt. Print. Off. 1976.
CO viii, 89 p.: ill. ; 25 cm.
SS NASA SP; 407.
SU Reusable space vehicles.
SA United States. National Aeronautics and Space Administration. NASA SP; 407.
NT S/N 033-000-00651-9.
NU GPO: NAS-1-21-407. DCN: TL875-L963-1976.
I.
FF PD: 1976. CP: dcu. LG: eng.
AN oc172519972.

FIGURE 3.8

Author:	Rotenberg, Manuel.
Title:	The 3-j and 6-j symbols (by) Manuel Rotenberg (and others) Cambridge, Technology Press, Massachusetts Institute of Technology, 1959. 498 p. 29 cm.
Notes:	Includes bibliography.
Subjects:	Angular momentum (Nuclear physics) Nuclear physics -- Tables, etc.
Call numbers:	UCI Phy Sci QC174.5 .R65 Oversize (CU-I-PS) UCR Phy Sci QC174.5 .R65 Ref (CU-RivP) UCSD S & E QC174.5 .R843 (CU-SSe)

FIGURE 3.9

```
 2 040          NCS c NCS d m.c. d OCL
 3 092          347.73064 b W726½
 4 049          OCCL
 5 100 10       Williams, Robert F., e comp. w cn
 6 245 10       Legality of microfilm : b admissibility in evidence of
                microfilm records / c edited by Robert F. Williams.
 7 260 0        (Chicago) : b Cohasset Associates, c c1980-
 8 300          1 v. ; c 29 cm.
 9 500          Loose-leaf for updating.
10 650 0        Microfilms.
11 650 0        Evidence, Documentary z United States.
12 650 0        Documents on microfilm x Law and legislation z United States.
```

FIGURE 3.10

LO DRBB.
HT115-M38.
AE Lopez, Robert Sabatino, 1910-.
TI The Medieval city / edited by Harry A. Miskimin, David Herlihy, A. L. Udovitch.
IM New Haven: Yale University Press, 1977.
CO x, 345 p.; 25 cm.
SU Lopez, Robert Sabatino, 1910-. Cities and towns, Medieval Addresses, essays, lectures.
NT Includes index. "Publications of Robert S. Lopez": p. 329-334. The Italian city: Herlihy, D. Family and property in Renaissance Florence. Krekic, B. Four Florentine commercial companies in Dubrovnik (Ragusa) in the first half of the fourteenth century. Lane, F. C. The first infidelities of the Venetian lire. Cipolla, C. M. A plague doctor. Kedar, B. Z. The Genoese notaries of 1382. Hughes, D. O. Kinsmen and neighbors in Medieval Genoa. Peters, E. Pars, parte: Dante and an urban contribution to political thought. --The Eastern city: Udovitch, A. L. A tale of two cities. Goitein, S. D. A mansion in Fustat. Prawer, J. Crusader cities. Teall, J. Byzantine urbanism in the military handbooks. --The Northern city: Miskimin, H. A. The legacies of London. Munro, J. Industrial protectionism in Medieval Flanders. Strayer, J. R. The costs and profits of war. Hoffmann, R. C. Wrocaw citizens as rural landholders. Cohen, S. The earliest Scandinavian towns. FOR FULL INFORMATION SEE AUTHOR CARD.
NU LC: HT115-M38. DDC: 301-36-3-094.
LCCN: 77076302. ISBN: 0300020813.
FF PD: 1977. CP: ctu. LG: eng.
AN oc173071420.

FIGURE 3.11

1
TI Relativity, quanta, and cosmology in the development of the scientific thought of Albert Einstein / director, Mario Pentaleo; editor, Francesco De Finis.
SU Einstein, Albert, 1879-1955. Cosmology.

2
TI Albert Einstein's theory of general relativity / edited by Gerald E. Tauber.
SU Einstein, Albert, 1879-1955. Physicists Biography.

3
TI The ethereal aether; a history of the Michelson-Morley-Miller aether-drift experiments, 1880-1930, by Loyd S. Swenson, Jr. With a foreword by Gerald Holton.
SU Relativity (Physics) History.

However, a recent study by OCLC of operational OPACs indicates that not all OPACs will support the sophisticated information retrieval techniques discussed above (9). For those OPACs that use phrase title. Also, this method assumes the availability of online access with sophisticated retrieval capabilities.

III. Augmentation of LCSH

Online search capability over MARC record fields will improve subject retrieval but will not solve the problems associated with the sparseness of subject headings (an average of 1.7 per title) and lack of indexing exhaustivity. A study by Pauline Atherton Cochrane, carried out under a Council on Library Resources grant, showed that augmentation of LCSH with terms taken from the table of contents and body of the work can effectively improve online subject access to monographic materials (11).

The price paid for this improved access is a doubling in the length of the MARC record and an investment of 12 minutes of additional subject cataloging time per title. Also, this method assumes the availability of online access with sophisticated retrieval capabilities.

Another approach to augmentation of LCSH is to identify distinguished subsets of a library collection for which enhanced subject access is most important.

An Iowa State University project identified techniques for augmenting subject access to reference collection materials (12). Reference materials are approached primarily by patrons with information needs, as opposed to known-item searches, and often contain detailed tabular, statistical, and directory information. The project developed a thesaurus for reference materials and studied the streamlining of data entry in order to minimize the increase in record size. A database covering some 6,000 reference titles was

created, from which a printed and microform index was produced. More recently, Lockheed has announced the availability of Superindex, a database containing the back-of-the-book indexes of some 500 reference works, including the annual review series.

Brett Butler has suggested that this idea can be carried a step further by identifying a subset of the titles published annually to which additional subject access points can be attached (13).

In any augmentation of LCSH, there are a number of possible approaches:

1. Additional LCSH can be assigned to titles or selective titles. In fact, the assignment by LC of more subject headings per title would improve subject access;

2. Uncontrolled terms from the table of contents and work itself can be entered into the MARC record;

3. Descriptors from subject specific thesauri can be assigned. Niehoff has done some work on connecting the heterogeneous descriptors from different thesauri using a vocabulary switching mechanism (14).

4. Subject headings from alternative schemes such as the Hennepin County subject headings list or PRECIS system can be assigned.

IV. Subject authority control

The production by LC of timely and complete machine-readable subject authority files (including non-print headings) would lead to improved subject access. Integration of this information into OPACs would facilitate elimination of obsolete headings and allow the automatic replacement of a user-entered variant heading form with the authoritive form of the subject

heading. Online subject authority control would provide a systematic syndetic structure and increase the entry vocabulary for user searching.

There are other software-driven enhancements to LCSH that could be examined. These include 1) reformatting LCSH into a thesaurus-style arrangement, with BT (broader term), NT (narrower term), and RT (related term) indicators, and 2) linking LCSH with the LC classification schedule terms via the 053 field in subject authority records.

V. Bibliography

1. Freedman, Maurice J.; Malinconico, Michael S. editors; The Nature and Future of the Catalog. LITA's 1975 and 1977 Institutes on the Catalog. Phoenix, Az: Oryx Press; 20-33, 89-101, 120-121, 225-256; 1979.

2. Kirtland, Monika; Cochrane, Pauline; Critical Views of LCSH- A Bibliographic and Bibliometric Essay. Cataloging and Classification Quarterly. 1:71-94; 1982.

3. Gorman, Michael; How the Machine May Yet Save LCSH. American Libraries. 11:557-558; 1980 October.

4. Mischo, William; LCSH: A Review of the Problems, and Prospects for Improved Subject Access. Cataloging and Classification Quarterly; 1:105-124; 1982.

5. Besant, Larry; Users of Public Access Catalogs Want Sophisticated Subject Access. American Libraries. 14:160; 1982 March.

6. Atherton, Pauline; Catalog Users' Access from the Researcher's Viewpoint: Past and Present Research Which Could Affect Library Catalog Design. Closing the Catalog. Phoenix, Az: Oryx Press; 107-115; 1980.

7. Weintraub, D. Kathryn; The Essential or Desiderata of the Bibliographic Record as Discovered by Research. Library Resources and Technical Services. 23:291-305; 1979 Fall.

8. Kaske, Neal K.; Sanders, Nancy P.; Online Subject Access: The Human Side of the Problem. RQ; 20:56-58; 1980 Fall.

9. Hildreth, Charles R.; Online Public Access Catalogs: The User Interface. Dublin, OH; OCLC: 1982.

10. Mischo, William H.; Technical Report on a Subject Retrieval Function for the Online Union Catalog. Dublin, OH; OCLC Report OCLC/DD/Tr-81/4; 1981 November.

11. Atherton, Pauline; Books are for Use: Final Report of the Subject Access Project. Syracuse, NY: School of Information Studies: 1978 February.

12. Mischo, William H.; Expanded Access to Reference Collection Materials. Journal of Library Automation. 122:338-354; 1979 December.

13. Butler, Brett; Book Indexing Group. Menlo Park, CA; Information Access Corporation; no date.

14. Niehoff, Robert; The Optimization and User of Automated Subject Switching for Better Retrieval. Communicating Information. Proceedings of the 43rd ASIS Annual Meeting. 17:397-400; 1980.

* * * * * * *

Discussion

In reaction to Mischo's comment that many online public access catalogs do not allow use of scphisticated search techniques, and consequently an excessive number of postings often result, one participant suggested that the problem is minimal if only a few false drops occur. In situations where it costs $.10 to examine each hit, however, it is important to have appropriate search capabilities to limit the search. Prompts on the public terminal that suggest ways of limiting any search that produces too many hits are important and valuable. Online catalogs that have two levels of use--one more sophisicated, and the other less sophisticated, with more prompts to help the user--are very helpful.

If augmentation of records is done, items with subject heading subdivisions "addresses, essays and lectures" may be the most promising ones to augment first; for many such works the only access now available is by searching the contents notes.

3.4 DISCUSSION PAPER 3

WORD, PHRASE AND TERM (DESCRIPTOR) SEARCHING

by Elaine Svenonius

I. Definitions

Word: ("keyword"), a sequence of characters bounded at either end by a space or its equivalent (period, comma, question mark). Word searching can be performed on the various fields of a record to retrieve bibliographic citations, full text or nonbibliographic data. Word searching can also be performed on indexes, thesauri and the descriptor fields of records to retrieve the phrases in which they are embedded (as part of a two step search).

Phrase: a multi-word expression occurring in running text or in descriptors. The syntax governing the formation of multi-word descriptors can be natural (ordinary language syntax) or artificial, i.e., special to a given index language, e.g., LCSH or PRECIS. It can be a pre- or post-coordinate syntax. Phrase searches are used to retrieve citations, full text or nonbibliographic data. Conceivably they could be used to retrieve larger phrases in which they are embedded.

Term: (synonymous with "index term" or "descriptor"; since "term" is ambiguous, meaning "search term" or "index term", the expression "descriptor" will be used here.) A descriptor is a well-formed expression in a controlled vocabulary. Descriptors may include words (uniterms) or phrases (Library of Congress Subject Headings, PRECIS strings, etc.), any expression constructed

according to the syntax rules of a given index language. Descriptor searching is used to retrieve citations, full text or nonbibliographic data.

II. Issues

The following are some issues to be considered in terms of effectiveness gained per dollar spent. They should be considered not in isolation but with respect to other variables, such as:

1. State-of-the-art and future technology (hardware and software)

2. Subject discipline(s) of the data base

3. Database size

4. Database type: nonbibliographic or bibliographic (citation or full text; book or article length) or a combination thereof.

Issue 1: Are descriptors necessary?

A difficulty with research to-date on the issue of keyword vs. descriptor searching is that the question is addressed so globally that it masks several more specific questions relating to index language design, e.g.:

1.1 Under what conditions does keyword searching improve precision? recall? (This relates to Issue 3.)

1.2 To what extent do the words in bibliographic records (as they are constructed now or as they may be enriched) signify what a book or article is about?

1.3 To what extent does a post-coordinate syntax suffice in phrase searching to achieve desired precision (cf., the use of Boolean and proximity operators with the use of a precoordinate syntax, such as is given in LCSH or PRECIS). (This relates to 2.13.)

1.4 To what extent does a user's search vocabulary match the searchable vocabulary in bibliographic records?

Issue 2: Given that descriptors are necessary, what should they look like, how should they be related and how should they be displayed online?

2.1 What should descriptors look like? This is a question of vocabulary and pre-coordinate syntax, i.e., what constitutes the well-formed expressions of an index language. It can be addressed in the form of several more specific questions:

2.11 How large and in what proportion to each other should the entry and descriptor portions of a controlled vocabulary be? (This relates to 2.21.)

2.12 Is a consistent pre-coordinate syntax needed in an online environment? (e.g., Should an attempt be made to normalize LCSH syntax?)

2.13 Assuming descriptor searching will be used in conjunction with keyword searching: to what extent should the pre-coordinate syntax of an index language govern the combining and fracturing of words?

2.14 Assuming a descriptor vocabulary must include phrases as well as words: to what extent can descriptors be derived automatically? In particular, what advantages, if any would the automatic derivation of phrases have over proximity and string searching techniques?

2.2 How should descriptors be related? This is a question of semantics: what meaning relationships should the descriptors of an index language bear to each other?

2.21 Are synonymy relationships necessary? In free-text searching can the online user be relied upon to be aware of synonymous expressions and to link them with Boolean OR's? In controlled vocabulary searching, how large (rich) an entry vocabulary is needed?

2.22 Are hierarchical relationships necessary, and if so, how rigorously are they to be defined? This question might be considered in light of whether the index language is to be used in conjunction with a classification scheme (i.e., LCSH plus DDC?) or as part of a switching structure.

2.23 Are related-term relationships necessary, and if so, how are they to be defined? Answers here might consider whether related descriptors might be got easily enough by browsing the descriptor fields of retrieved relevant citations.

2.3 How are descriptors to be displayed in an online environment? Several possibilities suggest themselves:

2.31 Not at all; the user can make use of a manual thesaurus, or guess or get ideas about useful descriptors from retrieved bibliographic records.

2.32 In alphabetic sequence, as is done by the NEIGHBOR or EXPAND commands.

2.33 Grouped by keyword. For instance, someone interested in education can ask to see all subject headings in which the word "education" occurs. These might be listed alphabetically or rotated as done in KWIC indexes.

2.34 Conjoined with associative and/or hierarchical information in the context of a classification or arrowgraph.

Issue 3: Are descriptors enough?

In particular are descriptors, combined with keywords sufficient to achieve desired precision and recall levels? Need they be augmented by classificatory devices such as concept codes (or the use of the EXPLODE command) or the feature headings of a classification scheme? Need they be combined with enhancements such as abstracts, codebooks, tables of contents, etc.?

Issue 4: What is a desirable depth (exhaustivity) of indexing?

In the context of keyword searching this is the question of how much text, and what kind, should be searchable. In the context of descriptor searching it is the policy question of how many descriptors should be assigned to a document.

* * * * * * *

Discussion

One participant asked if some of the concern about excessive postings isn't superfluous; in many cases the system is just doing what it is asked to do. For example, a large posting is automatically expected if one asks for items on "education" in the ERIC database; users who get such large postings learn more about how the system operates, even if they are mildly inconvenienced at the time. Systems have to be sophisticated enough that both novices and professionals can use them, and that high precision or high recall can be selected for.

3.5 DISCUSSION PAPER 4

CLASSIFICATION AS AN ONLINE SUBJECT ACCESS TOOL
Challenge and Opportunity

by Pauline Cochrane

Projects as early as 1965-66 demonstrated that searching a classification schedule online could result in a helpful array of related items during the browsing portion of a search and successful results could be obtained from selecting class numbers and using them as search "terms". Projects using the MARC Pilot Project tapes demonstrated that DDC and LCC class numbers, used in conjunction with the Library of Congress Subject Headings (LCSH) and title keywords, could bring recall up to and over 90%, when no subject access field could do so well alone. Operational online public access catalogs (OPACs) studied by Hildreth in 1981 showed half of them with a "call number" search

capability, sometimes truncated. If all this evidence points toward subject access via classification, why are we asking the question? In my opinion it is because the library world does not seem ready for classification to be used as an online subject access tool. To be ready implies more than a mere list of call numbers from MARC records which can be scanned online. To be ready means re-examining the work of classification and the impact of online access on that process; it means viewing the role of classification efforts from the online searcher's point of view and reviewing what can be done to improve his/her satisfactory searching and results. My analysis of this challenge parallels the work Karen Markey and I did for Central ERIC when they asked us to review the impact of online searching (especially free text searching) on the ERIC database (ED 180432). We had to separate suggestions for improvements into three groups, depending on who the initiator for improvements would be. As in that case, I think we must review the role and efforts of:

1. Classification makers and maintainers (DDC and LCC);

2. OPAC designers (at LC, RLG, WLN, CLSI, and a myriad of libraries, including NLM, OSU, Northwestern, etc.); and

3. OPAC users as represented by library staff (public and technical), and typical "end" users.

To be ready, as I said earlier, means more than listing LC call numbers and DDC class numbers from the MARC records. It means linking class numbers with subject headings (something promised as early as 1927 in the LCSH introduction). Several persons outside LC have tried to do this--to mention only a few: Mannheimer, Williams, and Daily; Nancy Olson; Bowker in their Subject Authorities. Because few online searchers will care to check all the LC schedules to learn what their favorite class numbers mean, or the Dewey

numbers, either, some attempt will have to be made to "translate" these numbers and develop table look-ups for online display, perhaps similar to the LC Classification Outline.

The comparable effort for DDC would be a publication of all DDC summaries. Any effort like this, of course, must be viewed as a publication for the online searcher, or a "table" for the online system to display when needed. If it is not viewed as a user's tool, we will create something only useful for the classifiers and catalogers, the intended audience for all of the above-mentioned attempts to combine subject headings and class numbers.

Issue 1: Shelf arrangement problems have served as a brake on revising classification schedules or on demonstrating their utility as a subject access tool. Can we divorce shelf arrangement as a process from classification as a subject access process so that we can create useful, systematic browsing displays online which would provide a helpful order of items, avoiding the problem of the inherent order of alphabets?

Both Michael Gorman (American Libraries, September 1981, p. 498-9) and Nancy J. Williamson (Library Resources & Technical Services, April/June 1982, 122+) have addressed this issue briefly.

The preliminary results of the CLR-sponsored OPAC User Evaluation studies have shown that users need assistance when their search results are either too few or too many. They also request viewing "terms related to their search". System designers are perturbed about response time when terms which are "too common" are used. Can class numbers in MARC record classification schedules in auxiliary online files help in this area? Retrieval system designers will have to be creative in this area, more so than presently evidenced. Perhaps some developmental effort needs to be supported again, as was done in 1965-66 (AIP Project AUDACIOUS, funded by NSF). Beyond using the EXPLODE command via MeSH's tree structures, a quasi-classification, there do not seem to be any

ingenious uses of classification to broaden or narrow searches, or to improve response time by translating common terms entered into a system message to prompt the user to narrow their search, etc.

Issue 2. Can a combination of OPAC designers, researchers, and classification owners and maintainers come up with some ingenious uses of class numbers to improve response time online, to guide users to better search strategies, etc.?

Issue 3. What useful links can be forged between LCSH and LCC or DDC which will be helpful online in various OPACs?

This is an issue related to both Issue #1 and #2 because most people have missed the value of these links. NLM has not, interestingly enough, in that the schedule they use for CATLINE (Class W) is indexed using MeSH terms, as much as possible. Has that day come for LCC? If it did, what would we do with the result online? The data in the Bowker publication could be of some use for such a study, but taken alone, it can not yet show what the potential is online. Even NLM in writing the specifications for Medlars III has not incorporated any use of their Class W Schedule online, even though it is maintained in machine-readable form.

Subject Authority control can be viewed as a problem for one system of subject headings or a single classification system, or it can be viewed as a problem of users who search in multiple files, each with their own unique vocabularies or classification systems. If viewed as the later, something needs to be done to integrate, if not make compatible, the various systems which might be searched. All through the 1960's attempts at compatability were made, but we have not yet seen any results in this area which have changed the life of online searchers. Retrieval system vendors like SDC, DIALOG, and BRS have attempted multiple-database vocabulary indexes online, but these are the lowest common denominator, simply a merger of lists of terms from database records, with no attempt to show the syndetic structure of each

vocabulary or to group these terms into broad related groups. The Integrated Energy Vocabulary and the Battelle Switching Vocabulary System are examples of things to come. The application of BSO (Broad System of Ordering; a UNISIST project) is another way to proceed. Which way will lead to the biggest payoff from the least investment?

Issue 4. Are there automatic means for achieving online switching between subject vocabulary and classification systems? Does an effort like BSO have to be imposed before multiple files can be searched adequately online with the least user effort? Is a transparent translation from a user's search terms to the system vocabularies in an OPAC feasible and practicable?

All of the above discussion assumes that we would not get bogged down by the idiosyncracies of our present systems which have tried to accommodate shelf arrangement, format considerations, and interpolation problems. In other words, if we start off assuming classification can serve a useful purpose online, what might these purposes be and how can we get there from existing records, existing schedules, and existing systems?

* * * * * * *

Discussion

One participant mentioned that the authority format allows the use of notes that could help the user know where to look for what he is seeking. Participants in some of the focused interviews conducted as part of the online public access catalog evaluation project have talked about the idea of a "knowledge tree," which could help a user broaden or narrow a search.

* * * * * * *

CHAPTER 4

DEVELOPING RECOMMENDATIONS

During the Tuesday afternoon session the group was divided into four subgroups to further discuss the morning topics (Chapter 3) and to develop recommendations for action. Subgroups were instructed to select ideas from the entire range covered (and go beyond if necessary), to identify activities that could be done in both the near-term and long-term to help improve subject access for the user, and to present those suggested activities as recommendations. Discussion leaders from the morning session served as discussion leaders of the subgroups, and one person in each subgroup was asked to take notes of the discussion and report the sense and substance of the subgroup's deliberations when the entire group re-convened in the evening. Section 4.1 identifies the membership of the subgroups.

The Tuesday evening session was devoted to subgroup reports. Sections 4.2, 4.3, 4.4 and 4.5 summarize the reports of the subgroups, indicate any discussion that occurred when the reports were given, and list the recommendations made. A total of 40 recommendations were made, and one group (C) proposed 7 basic assumptions underlying their recommendations.

4.1 THE SUBGROUPS

Participants were divided into four subgroups in order to stimulate discussion and give all participants an opportunity to contribute. Assignments to groups were made by CLR staff; the intent was to insure a diversity of backgrounds within each subgroup. Library administrators and

policy makers, researchers, and staff from network services, database services, and indexing and abstracting services were mixed together in the subgroups.

The subgroup strategy seemed to work well. Discussion was lively, and progress was made despite the broad nature of the task. Approaches taken by different subgroups varied, but common threads frequently ran through their recommendations. Reactions of the participants to the small group work was basically positive, although it was suggested that the groups might have been a bit more effective if the discussion leaders had received more direction and preparation time.

Discussion leaders are designated (L) in the listing below; reporters are designated (R). For subgroup B two speakers from the morning session shared the discussion leader responsibilities. The subgroups were:

Subgroup A

David Bishop
Charles Bourne
Judy Herschman
Neal Kaske (R)
Lucia Rather (L)

Subgroup B

Alan Benenfeld (R)
Brett Butler
Tina Kass
Jim McDonald
Mary K. Pietris (L)
Elaine Svenonius (L)

Subgroup C

Jeff Griffith
Carol Mandel
Dave McCarn
Bill Mischo (L)
Martin Runkle
Velma Veneziano (R)

Subgroup D

Pauline Cochrane (L)
Tamas Doszkocs
Doug Ferguson
Donald Hawkins
Pat Molholt (R)
Jennifer Younger

4.2 RECOMMENDATIONS OF SUBGROUP A

Subgroup A (Bishop, Bourne, Herschman, Kaske, and Rather) first divided the issues into three categories: (1) systems issues, which included user issues (such as ease of use), free vs. controlled vocabluary, Boolean operators, search strategies, multiple modes, and transparency; (2) subject heading issues, which included those issues related to controlled vocabulary and classification; and (3) database issues, which included enhancement, cost, storage, and the handling of multiple class numbers. The subgroup devoted most attention to categories 2 and 3, and made eight recommendations:

1. The Council on Library Resources should facilitate communication between and among systems developers, managers, and users, recognizing that there will be independent development. Possible actions would include bringing together people and ideas in various ways, perhaps by (a) collecting and disseminating information about systems; (b) sponsoring annual conferences with published proceedings; (c) publishing works which describe new trends and new systems; and (d) funding product evaluation studies.

2. CLR should continue funding research on online public access catalog systems. Particular emphasis should be on transaction analysis and end user studies.

3. Promote work on database enhancements. Create a mechanism for authorized special interest groups (or indexing and abstracting groups) to augment existing records, perhaps by adding specialized thesaurus terms (e.g., MeSH) to records, or table of contents data, etc. LC would still create the bibliographic record, but source material could be provided by specialized organizations.

(Discussion: One participant questioned whether it is a technical, legal or political problem if one wants to augment the record once it leaves LC. At that point, the only way to add to the record is by creating your own.

Kaske mentioned OCLC will have a project "Enhance" later this year or next year which will allow people to change the master record on OCLC. Several persons commented they did not want to hold up a record for three years while it gets enhanced; perhaps LC could release the record, and update it later in enhanced form.)

4. Create an online subject authority file immediately, with a mechanism for frequent updates, to be distributed in machine-readable, LC-MARC authorities format.

5. Set up a mechanism for sending suggested synonyms and see references to LC, perhaps using electronic mail. Reference librarians and catalogers are logical contributors.

6. Eliminate obsolete terminology from LCSH.

7. Immediately begin a study of the correlation between see also and see also from references with broader terms (BT) and narrower terms (NT), to see what hierarchical relationships exist. The chief question is this: Could a program be written that would automatically convert see also and see also from references in LCSH to brmader terms and narrower terms? In what percentage of cases would such an algorithm work successfully? How much, and what kind of, manual intervention would be needed?

8. Create and distribute a detailed LC Classification Schedule with scope notes in machine-readable form, and update it on a regular basis. This would facilitate a number of things that need to be done.

4.3 RECOMMENDATIONS OF SUBGROUP B

Subgroup B (Benenfeld, Butler, Kass, McDonald, Pietris, and Svenonius) made the following 14 recommendations:

1. In principle, a cooperative subject authority control system at the national (or even international) level is desirable. Studies are necessary to determine how best to develop and operate one.

2. Parallel development of databases from existing printed tools is desirable; tools involved include LCSH, Sears subject headings, specialized thesauri, and LC, Dewey and specialized classification schedules. Such databases could be very useful in online public access catalogs. Subject heading databases would make it possible to do the following: (a) display all authorized elements and provenance of the LCSH in one place, along with elements of class codes; (b) display all LCSH strings (main headings and subdivisions) as used in practice in records; (c) link subject authorities to bibliographic records; (d) specify usage elements, such as scope notes; and (e) display syndetics (see references, with provenance, and see also references).

3. Test the utility of having classification schedules online.

4. Develop mechanisms to help users enter authority systems and work their way through such systems to get to bibliographic records. Such aids should be transparent to the user.

5. LCSH will be the basis for any major effort in subject authority control. Production and timely distribution of LCSH for system support is critical, and such distribution should respond to the needs of institutions outside LC.

6. Explore further whether or not LCSH should be enhanced.

7. Consider developmental work that relates parts of LCSH to specialized thesauri.

8. Explore further the impact on LC (i.e., staff time, resources needed, etc.) of proposed changes that could be made in LCSH for the benefit of online public access catalogs.

9. Create friendly interfaces that promote subject access by providing intellectual linkages from a broad entry vocabulary.

10. Explore ways in which huge databases can be subdivided into more manageable file sizes, and how such fracturing of databases could affect subject access. Can large files best be divided by discipline? Or by institution, based on holdings?

11. Conduct research on the feasibility of, and advantages of, using classification schedules as online subject access tools.

12. Promote research on the evaluation of interfaces to support subject access on different online public access catalog systems.

13. Conduct research on the depth of indexing of monographs. Questions to be addressed include: who could do it; how could it be done; how would it relate to depth of indexing of journal literature; what types of monographs would benefit most from in-depth indexing; and how could depth of indexing relate to results of user studies (e.g., what does the user find relative to what he/she is seeking?).

14. Make LCSH available in machine-readable form as soon as possible for use in online systems. The strings actually used in indexing (i.e., those that appear in bibliographic records and usually consist of a subject heading

modified by a subdivision) should also be available in online systems, and such strings should be linked with the records in the bibliographic file.

4.4 RECOMMENDATIONS OF SUBGROUP C

Subgroup C (Griffith, Mandel, McCarn, Mischo, Runkle, and Veneziano) worked from the following assumptions in arriving at their recommendations:

a. LCSH will be the basis of any controlled vocabulary.

b. It is too early to standardize a model of features and design of the "ideal" online public access catalog system.

c. Without additional funding, local libraries will continue to use LC records as available. The chances of getting voluntary enhancement of records on a large scale is slight.

d. The optimum tool for subject access is an online public access catalog that includes both (a) sophisticated search capacity (both free text and controlled vocabulary searching) and (b) holdings and availability information.

e. It is difficult now for most libraries to change headings retrospectively, but in ten years online public access catalogs will have authority control with global change capacity.

f. Local online public access catalogs may be sub-optimal. They may be adequate and valuable in the local setting, but they can be much more valuable if there is local access to more powerful systems for additional information.

g. In the future, journal articles may or may not be represented in online public access catalogs or in the large bibliographic databases offered

by the utilities. User populations are demanding access to journal literature in existing public access catalogs, but the problems of interfacing such varied databases suggest such developments are not imminent.

Subgroup C made the following nine recommendations:

1. Sponsor a project that will speed the availability of LCSH in machine-readable form with regular and timely updates.

2. Establish a communication mechanism by which information related to the development of online public access catalogs could be exchanged. Perhaps a clearinghouse should be established, or an electronic mail network. Demonstrations of systems could be sponsored at ALA conventions.

3. Establish a cooperative mechanism by which LCSH entry vocabulary can be expanded and improved. Rapid review of proposed entries is crucial. The mechanism could be along the lines of other cooperative ventures, such as the Name Authority Cooperative Project (NACO) or CONSER.

4. Establish a mechanism for integrating LCSH with other specialized thesauri and for facilitating switching between the two. This could involve incorporating specialized thesaurus terms into LCSH, and linking LCSH terms to the more detailed subject structure of the specialized thesaurus.

5. Study the feasibility of editing and converting to machine-readable form both the Dewey and the LC classification schedules. Descriptors should be added to each to increase their usefulness, and online browsing of schedules should be made possible.

6. Continue emphasis on the analysis of transaction logs, and in particular expand on the work being done with logs of real searches to determine the terminology and search strategies used by users.

7. Study the feasibility of editing LCSH toward a more systematic syndetic structure, particularly with respect to see also references.

8. Study the forms of LCSH headings to identify ways of making LCSH more consistent, and assess the impact of such changes on LC and the library community.

9. Review means by which more subject headings could be assigned per book, and better indexing could be provided for parts of books. Possible ways of doing the latter include: (a) prepare a format in which publishers could provide machine-readable tapes of CIP-type information (tables of contents, indexes, etc.) and establish incentives for participation; and (b) identify a subset of books on which augmentation techniques could be tried--conference papers and proceedings may be good candidates for such augmentation.

4.5 RECOMMENDATIONS OF SUBGROUP D

Subgroup D (Cochrane, Doszkocs, Ferguson, Hawkins, Molholt, Younger) made the following nine recommendations:

1. Test the utility of LC and Dewey Classification Schedules in online catalogs. Existing bibliographic records do not contain adequate subject information, but use of class numbers on bibliographic records can help improve subject access using the information already available.

2. Examine the effects on subject access of integrating and/or associating two or more thesauri or subject heading lists. The concept involves mapping vocabularies of specialized thesauri to LCSH. When enhanced indexing is needed, one could easily move from LCSH into the specialized thesaurus terms.

3. Promote in-depth indexing of major reference works, perhaps using the LC reference collection, and perform cost-benefit analysis of the results and analyze applications to other segments of collections. If one looks at literature as a whole, reference collections contain much identifiable, unique material. A pilot project at LC could help identify the cost of such added indexing and the benefits of it. The question of benefits vs. cost could then be answered based on actual data.

4. Produce and field test the use of micro-processors for diagnosing search behavior and for aiding in training searchers. Each of these activities is time-consuming, and is (or will be) necessary at a large number of institutions. Standardized programs could be developed for use in any library to record and analyze transaction logs and to train users in how to use the system, how to do Boolean searches, etc.

5. Study cost effective ways to add subject access points to bibliographic records. Possible areas to be investigated include OCR input of tables of contents, use of book publishers' tapes, etc. Such methods could reduce the inadequacies of subject indexing without requiring additional intellectual effort by the subject cataloger/indexer.

6. Improve LCSH and LC classification by using expert working groups. LC is unable to do any additional work in this area, but special interest groups (e.g., a core group of art libraries) could share the responsibility and help accomplish more that needs to be done. Such groups could help: (a) rationalize the structure and content of relevant parts of LCSH and the LC classification schedules; (b) expand the vocabulary and lead-in terminology of each; and (c) convert each to machine-readable form for standardized distribution to users of online catalogs and for staff use.

7. Improve MARC records by: (a) adding information to the fields that already exist; (b) adding fields that need to be added; and (c) processing existing fields using computers.

8. Use transaction log analysis as a technique for improving the currency of descriptors and the lead-in vocabulary of LCSH.

9. Analyze and report on variations in three aspects of online public catalogs: (a) commands being used; (b) field or index names; and (c) field or index contents. Currently there is much variation between systems; e.g., what constitutes an "author" field in one system may be significantly different from what constitutes an "author" field in another system. Such inconsistencies will create problems later if systems are linked.

* * * * * * *

The next chapter reports how the recommendations of the subgroups were reviewed and evaluated in arriving at a final set of recommendations and priorities.

* * * * * * *

CHAPTER 5

REFINING RECOMMENDATIONS AND SETTING PRIORITIES

This chapter describes in detail the course of discussions and the conclusions reached during the Wednesday morning session of the Subject Access Meeting. The purpose of that session was to review the recommendations of the four subgroups (as reported during the Tuesday evening session; Chapter 4), refine them to represent the consensus of the group, and establish priorities for action.

The group worked from a synthesis of the Tuesday evening assumptions and recommendations prepared by Jones and Russell. Duplicate recommendations were eliminated, similar ones merged, and related ones clustered into four categories: subject heading issues, classification issues, database issues, and general issues. Six working assumptions and 23 recommendations were on that list. Some editorial liberties were taken in preparing the synthesis, with the expectation that oversights or slights would be cleared up during the Wednesday morning discussions.

The task of the group was organized in two phases. The first two-thirds of the morning were devoted to a lively discussion of the working recommendations in the synthesis. The object was to modify the recommendations as necessary and assess the importance of each. During this phase most recommendations were modified, some merged, and others added; 16 remained when the meeting ended. The assumptions did not change.

During the rest of the morning, the group worked to (a) separate the remaining recommendations into short-term and long-term categories and

(b) assign priorities within each group. Short-term projects were defined as those that could be started and substantially completed within three years; seven were assigned to this category. Long-term projects were those that could be started anytime, but probably could not be completed within three years; three recommendations were assigned to this category. The remaining six recommendations were considered important by the group but were not assigned to either category.

In this chapter the working assumptions and recommendations from the Jones-Russell synthesis are given, followed by notes of the discussion related to each (including revisions made), and the final priorities established.

Because of the details reported in this chapter, the casual reader may wish to skip to the summary of the recommendations presented in chapter 6.

5.1 ASSUMPTIONS

The following six assumptions from the synthesis were accepted by the group.

1. The Library of Congress Subject Headings (LCSH) will be the basis for the controlled vocabulary in online public access catalogs.

2. It is too early to look for, or seek, standardization of features (including subject access strategies) among the many online public access catalogs available.

3. Local libraries are likely to use LC-MARC "as is" and will be unwilling to routinely enhance records.

4. The optimum subject search tool is the online public access catalog equipped with sophisticated search capabilities including natural language and controlled vocabulary searches.

5. Within ten years all online public access catalogs will be equipped with a global change capability that permits, for instance, a change of a single term in a controlling vocabulary to be reflected in every record using that term.

6. Less comprehensive online public access catalogs must have the capacity, when necessary, to search more comprehensive ones, through an appropriately designed procedure.

5.2 SUBJECT HEADING ISSUES

Working Recommendation 1. Distribute LCSH in machine-readable form in the LC-MARC authority format; provide for regular updates.

At the time of the meeting, the most up-to-date version of the LCSH file available in machine-readable form at LC was December 1978; a new edition updated through 1980 was in preparation. In order to act on this recommendation, changes for 1981 and 1982 would have to be added to the newest edition. A more serious limitation of the present file is its format, which is cumbersome but adequate for printing, but would require extensive revisions before it would be useful for online catalog purposes. Some members of the group said that no machine-readable file should be distributed until the file is almost "perfect" (i.e., both cleaned up and reformatted), but others said that it is so important to distribute a machine-readable file quickly that some imperfections in the initial product can be tolerated.

Participants assigned this recommendation a number 1 priority on the short-term project list. (Recommendation 21 also became a number 1 priority on that list; they were judged to be of equal importance.)

Working Recommendation 2. Establish a mechanism for reference librarians and catalogers to submit synonyms (see references) to LC for inclusion in LCSH.

This recommendation was merged with recommendations 7 and 16 to read as follows: "Develop a cooperative mechanism for a set of libraries to contribute subject headings to LC, and a rapid review process that enables LC to handle those contributions. Establish a mechanism for reference librarians and catalogers to submit synonyms (see references) for possible inclusion in LCSH. Use transaction log data to improve LCSH entry vocabulary."

The revised recommendation was assigned the number 2 priority on the short-term project list.

Working Recommendation 3. Eliminate obsolete terminology from LCSH.

Rather said that it is expensive and time consuming to change or delete obsolete terminology, but LC does make such changes frequently. Pietris added that LC currently has a backlog of approximately 1,000 obsolete terminology suggestions that need to be evaluated for possible action. Suggestions of obsolete terminology could be solicited from practicing librarians, but before doing so the possible impact of such an effort on LC should be weighed against the benefits. Also, there needs to be a place in the authorities format to accommodate obsolete headings. No priority was set for this recommendation.

Working Recommendation 4. Examine see also references to determine the present degree of hierarchical relationships within LCSH. Investigate the possibility that a computer program could recognize such relationships and convert LCSH to a hierarchical thesaurus.

This recommendation was merged with recommendations 6 and 8 to read: "Determine if LCSH can/should be converted to an hierarchical thesaurus structure. Examine see also references to determine the present degree of hierarchical relationships. If appropriate, alter LCSH to a more systematic structure, with an enhanced syndetic structure."

At least one participant predicted this could be facilitated in the near future by modifying existing software that handles records similar to those in LCSH. Another suggested this should be a number 1 priority, but others said it is more important to make the file available in machine-readable form first, then work on improving it.

The revised recommendation was assigned a number 4 priority on the short-term project list.

Working Recommendation 5. Preserve the provenance of the machine-readable LCSH over time.

The purpose of this recommendation is to guarantee that a history of changes that have been made in the file is maintained and made available to all users. Information about when a heading was first used, and changes that have occurred (such as mergers and separations) should be part of this history. Presently there is no mechanism in place at LC for preserving provenance; one needs to be developed. No priority was set for this recommendation.

Working Recommendation 6. Determine if LCSH should be, or can be, converted to an hierarchical thesaurus structure.

This recommendation was merged with recommendations 4 and 8. See 4 above.

Working Recommendation 7. Develop a cooperative mechanism for a set of libraries to contribute subject headings to LC, and a rapid review process for handling those contributions.

This recommendation was merged with recommendations 2 and 16. See 2 above.

Working Recommendation 8. Enhance the syndetic structure of LCSH to make the structure of the file more systematic.

This recommendation was merged with recommendations 4 and 6. See 4 above.

Working Recommendation 9. Assess the impact that imposing consistency on LCSH would have on LC and other institutions.

Limited discussion; no priority set.

5.3 CLASSIFICATION ISSUES

Working Recommendation 10. Distribute the LC and Dewey Decimal classification schedules (with scope notes) in machine-readable form, with regular updates.

There was strong support for having the LC and Dewey classification schedules available in machine-readable form for use in online public access catalogs to provide more powerful subject access. Several questions arose, however: How useful would it really be? How would it be used? What expense is involved, both for development and use? How much of each schedule should be available (e.g., abridged vs. full versions)? How useful would it be to provide links between the LC and Dewey schedules?

Several participants commented that Dewey offers more opportunity in the near future, since it is already in machine-readable form at Forest Press, the publisher of the Dewey classification schedules. Much could be learned from working with that file. Further work with the LC schedules could begin with the creation of a pilot machine-readable file of the HM-HX schedules, a recently issued part of the system which reflects the degree of complexity in the schedules as a whole. Transaction log analyses of results of early use of classification schedules in online catalogs could help answer a number of questions.

Since the LC and Dewey classification schedules are separate entities, this recommendation was split into two recommendations, 10A (Dewey) and 10B (LC), with the same wording as above, to be considered individually. Recommendation 10A was assigned the number 5 priority on the short-term list, and recommendation 10B was assigned the number 2 priority on the long-term list.

Working Recommendation 11. Test the utility of machine-readable files of the LC and Dewey classification schedules in online public access catalogs.

Because this recommendation is so closely related to recommendation 10, it was considered along with it; 11 was deleted as a separate recommendation.

5.4 DATABASE ISSUES

Working Recommendation 12. CLR should continue support of transaction log analyses and user studies.

Several participants remarked that transaction log analyses are very useful in helping to understand how different types of users use online public access catalogs and other online files. Information from such studies is valuable in improving systems, and should be widely disseminated. This phrase was added at the end of the recommendation: "including searching of online reference files."

The modified recommendation was assigned a number 3 priority on the short-term list of recommendations.

Working Recommendation 13. Create mechanisms that enable users to work through the authority structure to reach bibliographic records.

This recommendation was merged into recommendation 20 (see below).

Working Recommendation 14. Develop a means by which LCSH can be integrated with other thesauri, and ways of switching between them.

The switching mechanism mentioned here is essential to getting optimal use from any system. The phrase "and classification schedules," was added after the word "thesauri."

Participants showed a fair amount of interest in this recommendation, but no priority was set.

Working Recommendation 15. Produce microcomputer-based diagnostic and training programs for assisting users of online public access catalogs.

Recommendation 15 was reworded as follows: "Explore and evaluate means of training users of online public access catalogs."

The suggestion was made that it is just as important to encourage sharing of information between trainers (perhaps by holding one or a series of meetings) as it is to encourage such sharing between systems designers. This topic was recommended as one of the first that should be addressed in the "fostering communication" efforts of recommendation 21.

The modified recommendation was assigned the number 6 priority on the short-term list.

Working Recommendation 16. Use transaction log data to improve the entry vocabulary of LCSH.

This recommendation was merged with recommendations 2 and 7. See 2 above.

Working Recommendation 17. Enhance records with additional terms by using such means as special thesaurus terms and table of contents data (which

might be available from publishers along with other data in machine-readable form). Identify categories of materials to receive this treatment. Involve approved special interest groups in making such enhancements.

Wording of this recommendation was changed to: "Augment subject access in bibliographic records by using..." Recommendation 19 was merged with 17 to read:

"Assess the cost/benefit of in-depth indexing and other means of augmenting subject access in bibliographic records. Explore the possible use of approved special interest groups to help. Areas to investigate include the use of special thesaurus terms and the use of machine-readable tables of contents and other similar data from publishers. If in-depth indexing is feasible, determine the class(es) of materials which would benefit most from such indexing."

Users are demanding richer bibliographic records, and this recommendation represents a new approach that could help increase the effectiveness of current systems. The Library of Congress, other institutions, and various special interest groups all could contribute to such efforts.

The revised recommendation was assigned the number 1 priority on the long-term list.

Working Recommendation 18. Analyze command and record structure and access point indexes of online public access catalogs.

This analysis involves both (a) what the commands and functions are and (b) what is in the record that is being manipulated.

No priority was set for this recommendation.

Working Recommendation 19. Assess the cost/benefit of in-depth indexing of major reference works.

The phrase "of major reference works" was deleted from the statement, with the understanding that major reference works are among the set of materials that could benefit most from in-depth indexing, if such indexing proves to be cost-beneficial. This recommendation was merged with recommendation 17 (see above).

Working Recommendation 20. Design methods for moving the search from the user's language to the search language and, ultimately, to bibliographic records.

Recommendation 13 was merged into this recommendation. One point of discussion was whether the user needs to be aware of the transition. Some participants felt that the transition should be entirely transparent, done without the user knowing it. Others felt the user should be notified that the transition is taking place, and why--this would help the user understand how the system works. The search should be completed without delay, and without the need for added keying or effort by the user, however.

This recommendation was assigned the number 3 priority on the long-term list of projects.

5.5 GENERAL ISSUES

Working Recommendation 21. CLR should accept a role in fostering communication among system developers, library managers, and others, perhaps by sponsoring special conferences (to be followed by printed summaries) and product evaluations. One topic that might be considered is the factors to be considered in selecting systems.

Some participants suggested that this recommendation, as written, does not fully reflect the magnitude and proactive nature of the role that should be assumed by CLR. For that reason, it was rewritten and combined with the related recommendation 23, as follows:

"CLR should take a leadership role in facilitating communication among system developers, library managers, and others, and should develop effective communication mechanisms to support developments in online public access catalogs and related areas. Ways of doing this might include: sponsoring special conferences and publishing summaries, producing or sponsoring product evaluations, and evaluating factors important in system selection decisions."

Two topics that could benefit from early attention under this recommendation are the training of users of online public access catalogs and the design of methods for moving users from their language to the search language and to bibliographic records.

The revised recommendation was assigned a number 1 priority on the short-term list; recommendation 1 was also a number 1 priority on that list.

Working Recommendation 22. Investigate whether an online public access catalog should only show the holdings of one library, or of a set of libraries. Could the full MARC set, and other records, be included?

The group agreed that this is an important question, and one which might be addressed as part of a larger study. No priority was set.

Working Recommendation 23. Develop effective communication mechanisms to facilitate developments in online public access catalogs and in related areas.

This recommendation was merged with recommendation 21, above.

The Subject Access Meeting ended at this point.

* * * * * * *

The next chapter summarizes these recommendations in priority order, with revisions in wording and emphasis recommended by participants after the meeting.

* * * * * * *

CHAPTER 6

SUMMARY OF RECOMMENDATIONS

This chapter contains all the assumptions and recommendations resulting from the Subject Access Meeting. Participants in the meeting have reviewed the text of this summary at least twice, and some changes to wording as well as additions of material have been made as a result of their having had time to reflect on the meeting.

Recommendations are divided into three sets. Two sets are equivalent in importance, the distinction being that one set is labeled Short-Term and the other set Long-Term Projects/Issues. The distinction between short-term and long-term is arbitrary but agreed upon by all participants. Short-term projects are those that can be started and substantially completed within three years. Long-term projects are those that may be started in the near term but hold little prospect for substantial progress within three years and will require effort well beyond the three year period.

The recommendations in the third set are those of lower priority, but which may nonetheless lead to useful projects for enhancing subject access for the users of online public access catalogs and other bibliographic reference systems. It should be understood that the subject access strategies discussed in this document are meant to incorporate not only the online catalogs now being introduced in libraries, but also are intended to cover other databases of bibliographic citations regardless of their source or location.

6.1 ASSUMPTIONS

All recommended projects and noted issues that were identified during the Dublin meetings assume several things. Among those assumptions are the following:

1. The Library of Congress Subject Headings (LCSH) will be the basis for the controlled vocabulary in online public access catalogs.

2. It is too early to look for, or seek, standardization of features (including subject access strategies) among the many online public access catalogs available.

3. Local libraries are likely to use LC-MARC "as is" and will be unwilling to enhance records routinely.

4. The optimum subject search tool is the online public access catalog equipped with sophisticated search capabilities including natural language and controlled vocabulary searches.

5. Within ten years all online public access catalogs will be equipped with a global change capability that permits, for instance, a change of a single term in a controlling vocabulary to be reflected in every record using that term.

6. Less comprehensive online public access catalogs must have the capacity, when necessary, to search more comprehensive ones, through an appropriately designed procedure.

6.2 SHORT-TERM PROJECTS/ISSUES

1A. The Council on Library Resources should accept a leadership role in developing an effective means of communication among several sets of people involved in issues related to online public access catalogs generally, and subject access strategies specifically. The sets of people that might be brought together for a variety of reasons include systems designers, those responsible for training users of online public access catalogs, those responsible for the selection of online public access catalog systems, those interested in development of innovative subject access strategies, etc. The techniques that might be used to accomplish these goals include topic oriented conferences with printed summaries and prepared papers to key the discussions, support of efforts to evaluate various online public access catalog systems and products, identification of appropriate factors in the selection of systems, cost considerations, both developmental and operating, etc.

1B. Create and distribute in machine-readable form the Library of Congress Subject Headings (LCSH) in the LC-MARC authority format providing for current and regular updates. The new edition of the LCSH in machine-readable form has just recently been released (since the meeting was held) and covers headings added through 1980 only. If other organizations were to help LC implement a subject authority file service with current and periodic updates, it would be necessary to rekey all of the 1981 and 1982 additions and changes to the file. Once that is done a subscription service could begin.

Items 1A and 1B were considered to be the absolutely top priority actions to be taken. In order to achieve maximum effect, they should both be started simultaneously. Online catalogs are just now developing as important library tools. The fact that there is much development going on now underscores the time-sensitive nature of these two recommendations.

2. Develop and establish a mechanism for a set of libraries to contribute new subject headings to the Library of Congress. For a larger set

of libraries, establish a similar mechanism for reference librarians and catalogers to suggest see references for inclusion in LCSH. In both cases, there should be a rapid review and reporting mechanism for all suggestions. In the case of cross references, transaction logs from online public access catalogs should be used as a source of suggestions.

3. Funding agencies should continue to support transaction log analysis and subsequent studies of the way different categories of users use online public access catalogs and reference databases, including those provided by the commercial sector. Because of the fundamental nature of the results of such studies, efforts must be made to see that this information flows steadily to those developing new online access systems.

4. Edit the LCSH see also structure so that true hierarchical relationships are made explicit and both broader terms and narrower terms can be distinguished and retrieved. Evaluate the LC subject headings themselves to determine if they can be rearranged and displayed hierarchically and whether such a change would be useful to users of online catalogs and catalogers. If such a display is not feasible, major improvements can and should be made in the syndetic structure of LCSH.

5. Encourage Forest Press to authorize the design of a machine-readable format for the Dewey Decimal Classification Schedules, including schedule, text and indexes, and to distribute, with periodic updates, for search and display only, the resulting database. One aspect of the project to create the format is the need to define what portion of the system should be converted to machine-readable form for public access purposes--the abridged edition or the full version--and, generally, how the database might be used for public access purposes.

6. Explore, evaluate and promote a variety of ways to train people with different backgrounds and experience to use online public access catalogs and other online bibliographic reference systems. Attention will have to be paid to the fact that increasingly such access will be available outside the library environment, a circumstance which calls for different training and updating strategies.

6.3 LONG-TERM PROJECTS/ISSUES

1. Identify and evaluate ways to augment and enhance subject access in newly created bibliographic records. These strategies might include additional effort on the part of the Library of Congress as well as efforts contributed by other institutions and special interest groups. The techniques that should be explored include but are not limited to the following:

a. Provide additional LCSH headings in bibliographic records (they may be considered secondary headings and not generate printed cards).

b. Add terms from special thesauri and see that these enhanced records are made available through LC, though the terms would most likely be added by other institutions or special interest groups.

c. Add table-of-contents data to the content note field, or as unformatted material appended to the record. Explore how this data, along with other possible data like index information, might come in machine-readable form from cooperating publishers.

d. Identify special groups of material to receive special in-depth indexing, for example, reference material or conference procedings.

Any alternative record enhancement strategy should be pursued only after assessing the cost/benefit to be expected from such enhancements.

2. Establish the utility of the Library of Congress Classification Schedules (with scope notes)--or edited portions thereof--in machine-readable form for users of online public access catalogs and catalogers. If they prove useful, design a format for update, search and display and organize regular, periodic updates. It is expected that incorporating some portion of the schedules with notes and indexes will provide a more powerful subject access infrastructure for online public access catalogs.

3. Design a basic and transferable strategy for moving from a user's language through the system's language for the retrieval of bibliographic records. A fundamental design criterion is that users need not be aware of the translation, some arguing for an indication of the nature of that translation so the user becomes more aware of the power of the system being used.

6.4 OTHER PROJECTS/ISSUES

There were several other projects recommended by the participants during the Subject Access Meeting, but these projects failed to achieve a high priority rating in the consensus process. The group did, however, find them sufficiently commanding that they survived a group discussion as viable and useful projects. One should take note of these in the context of what needs to be done to enhance and improve subject access for the user of online public access catalogs.

The following six projects are not presented in any special order.

1. Develop a way to preserve the provenance, over time, of each entry in the machine-readable LCSH. This means that a history of all headings changed or deleted from the file would be preserved and noted in the file itself.

2. Eliminate obsolete terminology and heading style from LCSH. It is possible that this specific objective would become part of any one of several other projects included under short- and long-term projects/issues. This work might be accomplished by agencies outside of LC but approved by LC.

3. Assess the impact, in terms of cost and service enhancement, of imposing consistency of form and language on LCSH.

4. Develop the basic strategy for integrating LCSH with other thesauri and with classification schedules, and develop ways to switch among them all.

5. Prepare an analysis of the command and record structures and the indexes maintained in a wide variety of online public access catalogs.

6. Attempt to evaluate whether or not online public access catalogs need to be restricted to the holdings of one or a set of libraries. Could/should they contain the full MARC record set as well as other records, such as one or more article citation or reference databases? What would be the local impact of providing access to records for which holdings information was not available?

* * * * * * *

The following chapter outlines developments related to subject access that have taken place since the Subject Access Meeting was held.

* * * * * * *

CHAPTER 7

EPILOGUE

This chapter reports some of the subject access developments that have taken place since the Subject Access Meeting was held.

7.1 THE WEDNESDAY AFTERNOON MEETING

Immediately after the Subject Access Meeting, eight participants met with LC processing staff and CLR staff to discuss the possible implications for LC of the discussions that had just taken place, and how progress could be made on some of the recommendations that would involve LC. Among the topics discussed were: distribution of an updated machine-readable LCSH, contributions to LCSH by other libraries, LCSH hierarchy, and bibliographic augmentation. Appendix D is a summary of that session.

7.2 FOSTERING COMMUNICATION

One of the number 1 short-term recommendations encourages CLR to facilitate communication among system developers, library managers, and others involved in issues related to online public access catalogs in general, and subject access in particular. CLR has scheduled two related meetings in the near future:

(1) In December, 12 designers of online catalog systems will meet with 12 research library administrators and appropriate resource persons to discuss

issues related to the features and costs of online catalogs. Participants will represent national and research libraries, vendors of online catalog systems, and other organizations. A report of the meeting will be published.

(2) In January, public services librarians, systems people, and vendors will meet to review strategies for teaching online catalog users to use systems effectively. This meeting specifically evolved from short-term recomendation 6: "Explore, evaluate and promote a variety of ways to train people with different backgrounds and experience to use online public access catalogs and other online bibliographic reference systems." A report of the meeting will be published.

7.3 LCSH IN MACHINE-READABLE FORM

Short-term recommendation 1B calls for the creation and distribution of LCSH in machine-readable form, with regular updates. This recommendation was discussed at some length at the meeting, at the Wednesday afternoon session that followed, and in several conversations since. One suggestion that has been pursued is the possibility of an outside agency creating and maintaining an up-to-date machine-readable file, and preparing updates, until LC could take over the project. It is now evident that LC will not be able to cooperate with an outside agency in putting weekly updates to LCSH in machine-readable form because of the complications of reconciling files later. LC will, however, be able to make an annual cumulation available, but more frequent issuance of the cumulation will not be possible in the near future.

7.4 CONTRIBUTIONS TO LCSH BY OTHER LIBRARIES

Short-term recommendation 2 involves development of a mechanism whereby a set of libraries could contribute suggestions of new subject headings and see references to LC for possible incorporation into LCSH. At the time of the

June meeting, Pauline Cochrane was working at LC on a project funded by CLR to establish a routine procedure by which new lead-in vocabulary (i.e., see references) for LCSH could be recommended to LC. The procedure would facilitate the receipt of cross reference suggestions so they could easily be channeled into the weekly review process at LC.

That procedure is now in place at LC, and four selected libraries (University of California at Berkeley, Duke University, Harvard University, and the National Library of Canada) are sending suggestions to the Subject Catalog Division of LC. The suggestions are reviewed routinely, and are either passed on to the Weekly Editorial Review Committee or are rejected and returned to the originating library with an explanation. By October the first of the accepted suggestions had appeared in the Weekly List, and over 100 suggestions were under consideration. LC staff are reviewing the results of this project.

7.5 LCSH HIERARCHY

Short-term recommendation 4 involves evaluation of the LC subject headings to determine if they can be arranged and displayed hierarchically, and how useful such a change would be to users. Tony Peterson of the Art and Achitecture Thesaurus project (AAT) reports on the AAT experience with LCSH, and has a paper in press on the topic. The following statement was prepared by her and Pat Molholt to summarize their experience:

"The Art and Architecture Thesaurus project developed out of a need for a comprehensive controlled vocabulary for these fields. The original funding which identified this need was provided by the CLR. Since then the NEH, the A.W. Melion and the J.P. Getty Trust have given the funds to produce the first section, on architecture. Recently funding has been announced from the Getty Trust which will secure the completion of the project over the next 2 1/2 years.

"Because the AAT is based on existing subject lists, including LCSH, and because the use of terms from these lists is tracked through the hierarchical structure of the AAT, a unique opportunity is presented to study how LCSH might improve its current structure and terminology.

"Of the 8000 to 9000 terms which will probably make up AAT/architecture, about 70% will have been lacking in LCSH and added from other sources. Two-thirds of the LCSH terms will remain unchanged, while about one-third will be modified for consistency's sake and in order to meet ANSI thesaurus standards. The modifications include inverting the headings to natural language order, using LC see references as main terms, adding parenthetical modifiers, removing conjunctions and otherwise splitting up multiple word terms, and making nouns plural.

"In addition, the LCSH syndetic structure was seen to be so inconsistent as to preclude its automatic conversion to a system of broader, narrower, or related terms. The AAT project team lists the following conclusions concerning the restructuring of LCSH:

1. LCSH can be structured hierarchically if experts within subject areas can be assigned the task of doing so for their fields.

2. When structured hierarchically and by subject areas, the vocabulary can be considerably enriched by the addition of new subject terms and access points, i.e., see references.

3. It is not possible to assign broader and narrower terms automatically to LCSH using current see also and see from terminology. There is so little logic existing in the current structure that it should probably be closed and the cross reference structure allowed to develop conceptually through arranging the main terms hierarchically.

4. LCSH subheadings function inadequately as an indexing system and are often assigned superfluously. Any system of subdivisions must have clearly set out prescriptions for use, along the lines of the MeSH form and topical subdivision rules."

7.6 ENHANCING SUBJECT ACCESS

Long-term recommendation 1 calls for the identification and evaluation of ways to augment and enhance subject access in bibliographic records. CLR staff have been discussing with at least one potential grantee possible ways to evaluate cost effective ways of enhancing bibliographic records for the provision of subject access.

7.7 REACTIONS TO RECOMMENDATIONS

The summary of recommendations from the Subject Access Meeting, as presented in chapter 6, has been distributed to interested persons who requested it. Each requester was invited to comment on the summary. The following suggestions have been made by some of those reviewers:

1. It is important to stress "...the desirability of making the MARC 520 field (Summary) searchable in networks and other systems and that cataloguers at LC and elsewhere be encouraged to provide summaries and abstracts more often than is now done."

2. Short-term recommendation 3 (support transaction log analyses as one means of studying how users use systems) should be broadened to "...make explicit the need for a variety of studies of catalog use, including the transaction log approach but also other approaches."

3. The wording of the first sentence of long-term recommendation 1 (evaluate ways to enhance subject access in bibliographic records) "...implies a commitment to the status quo in bibliographic records, because it speaks only of the possibility of augmenting/enhancing them. There are, in fact, the additional possibilities of reducing them (unlikely, but possible) and of altering their nature. These possibilities ought to be recognized. They should, of course, be under the same cost/benefit qualifications as the augmentation approach."

4. "...I would suggest that explicit attention be given in these recommendations to the need for achieving better and quantitative understanding of the benefits of catalog service."

* * * * * * *

APPENDICES

APPENDIX A

AGENDA OF THE SUBJECT ACCESS MEETING

Dublin, Ohio, June 7 - 9, 1982

Monday, June 7, 1982

5:30-6:00 p.m. Cocktails

6:00-6:45 Dinner

7:00-8:30 Program Session I

Welcome to OCLC and Dublin

Introduction of Participants

Introduction to the Bibliographic Service Development Program (BSDP)

Review of past BSDP efforts to begin work in this area

Review of present BSDP activities in this area

Statement of the problem; identification of issues to be discussed:

--Controlled Access

hierarchical thesauri
other thesauri
classification

--Uncontrolled Access

free text
word term

Outline of the plan of action for the rest of the meeting

...continued...

Tuesday, June 8, 1982

8:15-9:00 a.m. Breakfast at OCLC

9:00-noon Program Session II

Presentations and Discussions:

Subject Access in Library of Congress Catalog Records.
Lucia Rather and Mary K. Pietris.

Affordable Enhancements to Bibliographic Records for Subject Access. Bill Mischo.

Word, Phrase and Term (Descriptor) Searching.
Elaine Svenonius.

Classification as an Online Subject Access Tool.
Pauline Cochrane.

Noon-1:00 p.m. Lunch

1:00-4:30 Program Session III

Work in groups to identify options for action

5:30-6:00 Cocktails at the hotel

6:00-6:45 Dinner

7:00-8:30 Program Session IV

Reports by group reporters and discussion.

...continued...

Wednesday, June 9, 1982

8:15-9:00 a.m. Breakfast at OCLC

9:00-noon Program Session V

Summary and Recommendations

--Identify steps to take
--Identify participants
--Define time scale
--Evaluate prospects for real improvement
--Set priorities for action
--Review costs/benefits of each action

Noon Major program concludes; lunch available at OCLC

* * * * * * *

1:00-4:30 p.m. Wednesday Afternoon Meeting: Special session with selected participants to discuss subject access and the Library of Congress: how LC can move ahead; how other organizations can help LC in this area; etc.

APPENDIX B

PARTICIPANTS IN THE SUBJECT ACCESS MEETING*

Dublin, Ohio, June 7 - 9, 1982

Alan Benenfeld
University Libraries
University of California
Los Angeles, CA 90024
213/825-6515

David Bishop
University of Georgia Libraries
Athens, GA 30601
404/542-2716

Charles Bourne
Dialog Information Services
3460 Hillview Avenue
Palo Alto, CA 94304
415/858-3775

Brett Butler
Information Access Corporation
404 Sixth Avneue
Menlo Park, CA 94087
415/367-7171

Pauline Cochrane
School of Information Studies
Syracuse University
Syracuse, NY 13210
315/423-2911

Tamas Doszkocs
National Library of Medicine
8600 Rockville Pike
Bethesda, MD 20209
301/496-6531

Douglas Ferguson
University Libraries
Stanford University
Stanford, CA 94035
415/497-9724

Jeff Griffith
Congressional Research Service
Library of Congress
Washington, D.C. 20540
202/287-6447

Donald Hawkins
Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974
201/582-6517

Judy Herschman
University Libraries
University of California, San Diego
La Jolla, CA 92093
714/452-3336

Neal Kaske
OCLC, Inc.
6565 Frantz Road
Dublin, OH
614/764-6000

Tina Kass
Research Libraries Group, Inc.
Jordan Quadrangle
Stanford, CA 94305
415/328-0920

Carol Mandel
Association of Research Libraries
1527 New Hampshire Avenue, N.W.
Washington, D.C. 20036
202/232-2466

Davis B. McCarn
Online Information International, Inc.
6455 Windmere Circle
Rockville, MD 20852
301/530-8588

Jim McDonald
OCLC, Inc.
6565 Frantz Road
Dublin, OH 43017
614/764-6000

William Mischo
Engineering Library
University of Illinois
Urbana, IL 61801
217/333-3576

Pat Molholt
Art & Architecture Thesaurus Project
Folsom Library
Rensselaer Polytechnic Institute
Troy, NY 12181
518/270-6677

Mary K. Pietris
Subject Cataloging Division
Library of Congress
Washington, D.C. 20540
202/287-5342

Lucia Rather
Processing Services
Library of Congress
Washington, D.C. 20540
202/287-5333

Martin Runkle
Joseph Regenstein Library
University of Chicago
Chicago, IL 60637
312/962-8744

Elaine Svenonius
Graduate School of Library &
Information Science
University of California
Los Angeles, CA 90024
213/825-7641

Velma Veneziano
Northwestern University Library
Evanston, IL 60201
312/492-7641

Jennifer Younger
Memorial Library
University of Wisconsin
Madison, WI 53706
608/262-4907

CLR Staff:

C. Lee Jones

Keith Russell

*Addresses and phone numbers current as of December 1982

APPENDIX C

DEFINITIONS OF "SUBJECT ACCESS"

Each participant was asked to bring to the Subject Access Meeting his or her definition of "subject access." The following definitions were distributed at the meeting and briefly discussed:

Subject access is the ability to discover what a book is about by searching relevant information systematically supplied.

Anonymous

Subject access includes the languages, aids, protocols, and mechanisms for analyzing items on the basis of their titles and contents, analyzing expressions of user need for items having a particular content, and relating the two analyses through a search and display process using a storage and retrieval system. Items may be bibliographic or non-bibliographic; items may be represented in a system by surrogate records or could be stored directly as text or data. The languages may use natural or controlled vocabulary, and aids in their application may display syndetic, hierarchical, or other relationships among terms. There may be rules for assembling and permuting multi-word strings, and for reducing terms to word-stems, word-keys, or codes. Text analysis, user expression analysis, and the search process may contain mechanisms for the parsing of text and search expressions and the manipulation of elements through logic, positional occurrence, or statistical associative techniques. The search process should include or be directly linked to similar mechanisms operating on words contained in titles and abstracts of items. The search process may be iterative and with displays of language aids and intermediate results, including records and text, for search refinement. Although subject access primarily concerns expressions of a topical, geographical, or chronological nature, these may be implied by author or corporate body names, in which case name access becomes a surrogate form of subject access.

Alan R. Benenfeld

Subject access is used to identify one or more bibliographic items and is based on the content of the work or works. Data used for subject access may be a single element or a number of elements used in combination.

Types of data elements used for subject access include:

1. Text taken from the work itself; such as title, index, table of contents, and all or a portion of the main text.

2. Text about the work not taken from the work itself; such as abstracts and free form descriptors.

3. Controlled vocabulary descriptors; such as LCSH and other subject heading systems.

4. Classification strings which describe the subject of a work and place it with like works and in relation to other works.

The number of items retrieved through subject access may be further reduced using factors such as date, language, physical format and content characteristics such as the presence of illustrations and bibliographies.

David F. Bishop

Subject access is the set of processes and techniques used in representation of a work so that its content may become known to one desiring the information therein without prior knowledge of the existence of the work, its authorship, or location.

Brett Butler

Subject access is the man-machine interface, plus underlying syndetic data and information structures, plus processing algorithms to enable retrieval and presentation of bibliographic/textual records in response to an (unrestricted) topical search query.

Putting it more simply, subject access means allowing the searcher to find pertinent items in response to a (free-form English language) topical search question.

Subject access, as defined above, need not be limited to computerized book catalogs, but is equally applicable to other machine-readable textual or textual/numerical files or data bases.

Tamas Doszkocs

Subject access is the process of retrieving a set of records from a database as the result of a requirement statement supplied by a person with an expressible subject request. The requirement statement is matched against records in the database and the statement may be successively modified so that the resulting set (including an empty set) is satisfactory to the individual

or the process is discontinued for other reasons. The process operates within the time and knowledge constraints of the individual and the economic constraints of the organization funding the retrieval system.

Douglas Ferguson

Subject access is the systems and procedures that allow the user to locate material that is about something or someone. Of all the means of access, subject access is uniquely user defined: only the user can decide whether the work found contains information about the topic as he or she defines it. Methods of subject access must therefore be designed to, and be evaluated on the degree to which they enable users to locate items that meet their conception of the topic.

Jeff Griffith

Subject access is the ability to search for terms in a database according to topic or discipline. The terms may appear in any topic or discipline-oriented fields in the database, including (but not limited to) title, abstract, descriptor, or subject heading. Fields without subject-oriented information, such as author name, author affiliation, language, etc., are excluded from subject access.

Donald T. Hawkins

Subject access is a method (or system) of obtaining relevant informational materials addressing a given topic of interest. This method of access provides entry through any common vocabulary. The system provides direct entry at any level in a subject's hierarchy. A hierarchical classification structure of terms is maintained and can be viewed by users, thus related terms are also reported to users.

Subject access plus (SAP) is the expansion of a subject access system which permits users to limit "relevant informational materials" by date, format, language, etc.

Neal Kaske

Subject access is a path to needed items that is either: (a) constructed by analyzing and describing by pre-determined notation the intellectual content of the item, or (b) constructed by making use of descriptors already present in the bibliographic record for the item or in the item itself.

More simply, subject access is a way to find what you want if you know what it's all about.

Tina Kass

Subject access means finding something that has been written about an idea that you are pursuing when you do not have a specific, or "known," item in mind. Subject access can be achieved through word-of-mouth (e.g., do you know of a book on...?), through browsing shelves or files, by searching bibliographies, indexes and data bases, or, theoretically, by searching a library catalog. Successful subject access systems move a searcher as painlessly and effectively as possible from his conceptualization of what is needed to the materials or citations available.

Carol Mandel

Subject access is defined as any means of identifying needed material relevant to a given subject. Subject access implies search access points to catalog records via subject headings (controlled or uncontrolled), free-text terms from selected data elements, including cross-references, or hierarchical structure access points based on subject or classification structure.

Subject access (improved): Access points will be expanded from the current average of 1.8 subject headings to include greatly expanded subject access from subject headings, titles, cross-references, authors, and other selected elements. Access points would include access to book chapters or sections.

Davis McCarn

Subject access is a tool that is easily used by library patrons to retrieve information about a subject. A subject access method should be able to interpret a user's paraphrased subject request into an internal search that will ultimately lead to the retrieval of associated bibliographic information or related subjects for the user to search.

Jim McDonald

Subject access to library bibliographic materials allows library users to (a) locate materials on a given topic as expressed by a user statement or query off "aboutness" and (b) locate items for which the exact title or author may not be known. Instruments of subject access include the library catalog, abstracting and indexing services, bibliographies, and other indexes. Subject

access systems should (a) provide an entry vocabulary sufficient to match the search requests brought by users and (b) provide library users with suggested alternate entry forms and related terms.

William Mischo

Subject access is identifying basic concepts of interest, and retrieving documents pertinent to those concepts. For the Art & Architecture Thesaurus (AAT), this implies retrieval based not only on information inherent in the bibliographic record or the whole document, but retrieval of that document or record as enhanced by controlled indexing.

Pat Molholt

Subject access is any means of locating material on a given topic. This includes a structured approach (including subject terms, classification numbers, and codes) and a non-structured approach utilizing words in the title, series, annotation, table of contents, or any other data included in a record that can be searched.

Lucia Rather

Subject access is a means or a mechanism for identifying sources that contain wanted information or for retrieving the information itself. Indexing of sources of information can be accomplished by the assignment of words, phrases, or numbers that describe the subject. The subject may also be revealed by the author or title, and by the text.

Martin Runkle

Subject access is access to a document by what it's about (its theme, topic). It's aboutness may be denoted by a well-formed expression in a controlled vocabulary or by natural language words or phrases. The definition can be made partially operational by regarding a word as a sequence of characters bounded at either end by a space; a phrase as a significant sequence of words (e.g., a sequence that occurs above a certain threshold frequency); an expression in a controlled vocabulary as enumerated in, or constructed according to the syntax of, an index language. What is regarded as a subject is in part determined by what can be expressed by one or another index language.

Elaine Svenonius

In order to reasonably discuss subject access, one must first determine what are short term objectives and what are long term objectives. At Northwestern University, our short term objective is to provide, in our online catalog, subject access which is at least as good as that supplied by the card catalog which the online catalog replaces. This means that a searcher, using one or more words as an access term, is guided to a list of alphabetically ordered entries, which may be either established headings or references to established headings. He must be able to browse, rapidly, forward and backward in this list and having found one or more established headings which appear to meet his requirements, to browse, rapidly through a list of works under those headings. If he finds a term in the list which is not an established heading, he must be able to determine what he should use, (i.e., a see reference). In the case of established headings, he must be able to determine what are related headings. Whether see references are specifically assigned or automatically generated from key words or phrases, it is important that they be controlled so the user is able to detect the basic underlying structure of the catalog. The long term objective should be then to improve the structure of the vocabulary, to increase the level of subject analysis of works, and to increase the number of references available for use.

Velma Veneziano

Subject access encompasses two major types of searches: (a) the search to find bibliographic resources and information about a particular person, organization or topic, without knowing what bibliographic items will be found, and (b) a search to find a known item when the information necessary to directly search for that item is not known. Recently there has been an indication that persons using the catalog are searching by subject at least as often as they are searching for known items. Subject access must provide the means for successful completion of both types of searches.

Jennifer Younger

---

* * * * * * *

APPENDIX D

THE WEDNESDAY AFTERNOON MEETING

Following the Subject Access Meeting eight of the participants were invited to meet with LC processing staff and CLR staff to review those portions of the discussions that had implications for LC.

Participants in that meeting were: David Bishop, Pauline Cochrane, Jeff Griffith, Carol Mandel, Bill Mischo, Elaine Svenonius, Velma Veneziano, Jennifer Younger, Mary K. Pietris, Lucia Rather, Lee Jones and Keith Russell.

This appendix is a brief summary of that meeting.

1. Distribution of LCSH in Machine-Readable Form.

The current computer file of LCSH was designed in 1969-1972 for printing the "red book." The file is used in batch mode, and is cumbersome because a large amount of editing is required each time the file is used to produce a product. The last copy of the file distributed to the public was the 1978 master database tape; records on that tape are in the MARC authority record format. Within LC the annual print updates for 1979 and 1980 are available, along with the first quarterly update for 1981. The 1980 master database will consist of the 1978 master database merged with the changes made during 1979 and 1980.

At this time, a new master database tape can be produced every year. At least five major projects at LC have a higher priority than the distribution of LCSH in machine-readable form would have, so it could be two or three years before a version with update tapes will be available instead of the present system of issuing the new master database tape every year. Changes in the authorities format and the distribution mechanism at LC contribute to this delay. Further, LCSH is unique enough that Name Authority File Service software cannot quickly be adapted to help with this project.

Discussion focused on possible interim steps that could be taken to make at least something available in the next six months; the sense of the Subject Access Meeting participants was that this needed to be done quickly and as a high priority. Perhaps an external agency could assume responsibility for updating it until LC could take over; that agency could do the work and return the product(s) to LC for distribution. Northwestern University, Blackwell-North America, and UTLAS were mentioned as three agencies that might be able to do the work. The file created by the agency would have to be reconciled

with the file at LC when LC took over the creation of the file, but that reconciliation would be part of the contract. Close coordination with LC would be essential.

Another way of getting the updated file ready for distribution would be for CLR or another funding agency to provide money to LC to add the necessary staff for doing the work they plan to do. This could speed up the production and distribution of a machine-readable LCSH.

As a result of this discussion, staff at both LC and CLR will pursue these ideas further.

2. Contributions to LCSH by Other Libraries.

LC has already agreed to cooperative projects with 3 organizations (the Government Printing Office, Harvard, and the University of Chicago) in which these organizations may contribute to LCSH. However, LC can only do so much with existing staff, and they are not yet ready to launch into a subject authority cooperative project patterned after the Name Authority Cooperative Project (NACO). Extra work by LC staff would be needed to train contributors and review work, and the review process for subject work would have to be more thorough than it is for contributed name authority records. At the present time, the LC staff can only handle the workload caused by their own catalogers proposing new and changed subject headings and references. If a cooperative effort became very large, a full-time coordinator would be needed.

Ways of spreading the cost for such work among a large number of libraries are necessary. The role of LC as a national library is important, yet LC cannot be expected to foot the bill for everything. CLR or another funding agency should consider funding an investigation of alternative ways LC could finance some of its national-interest activities for cost recovery. Once the value of many of LC's activities is proven, it should also be easier to justify expenditures within LC.

The experience ERIC has had in the cooperative modification of its thesaurus by 16 branches is relevant, and should be reviewed. That two-year vocabulary improvement and review project, which involved much communication and resulted in the 9th edition of the ERIC Thesaurus, has been completed.

3. LCSH Hierarchy.

Editing the LCSH so that true hierarchical relationships between terms are made explicit was another high-priority item on the short-term list of priorities coming from the Subject Access Meeting. The main reason for the high priority was to help the user work his/her way through the subject headings, but a secondary reason is to provide internal consistency within LCSH in preparation for the day other thesauri will be linked to it.

The necessity of actually going to a hierarchical structure was challenged by those who think that (a) if the purpose of the hierarchy is just to stimulate the user's imagination, then hierarchical relationships may not be so important, and (b) if the purpose is to take online advantage of what is in the "red book," we should not get locked into broader terms and narrower terms, since better approaches may be available before long. It is possible, for example, to just display all see and see also terms as related terms, and let the user sift through them for better terms to use; one problem, however, is that for some terms the number of related terms is staggering.

One view is that what is really needed is a machine-readable file that contains as much information as possible, including the hierarchical relationships. Each system using that file can make use of the hierarchy, or not, depending on local needs. The point of disagreement seems to be how to get the hierarchical file in the first place, and how soon it will be.

What needs to be done right now is a quick study to see if a computer algorithm could convert LCSH to a hierarchical structure at low cost. If 70 to 90% of the headings could be correctly converted by the algorithm, it would be worth doing; manual editing would complete the conversion. The testing of an algorithm could be done quickly on a subset of the 1980 tape, and if the full project is judged worthwhile, a contractor could be hired.

As a result of this discussion, LC will do a small sample to see how often the conversion by a computer algorithm would be acceptable. If the results are good, they will consider using such an algorithm. In the meantime, other research regarding future editing should continue.

4. LC Classification Online.

Investigation of the use of the LC classification schedules was a high priority on the long-term issues list. Use of LC classification online could help users as well as record creators, yet only half of the online catalogs Charles Hildreth analyzed in his study allow search by call number, and those searches are for the most part unsophisticated. When Forest Press converted Dewey to machine-readable form, Arthur D. Little, Inc., was hired to do the conversion; the LC classification project is also a likely candidate for contracting.

One approach would be to do a pilot study with part of the LC classification schedule, or with the outline of the schedule. NLM has some experience in this area--they have converted the W class schedule of the NLM classification to machine-readable form.

5. Bibliographic Augmentation.

Investigation of possible ways of enhancing bibliographic records to improve subject access was another high priority on the long-term list. Previous studies need to be reviewed, and new ones done to explore what is

possible with the online capabilities that exist today, and the capabilities that will be possible as systems become more powerful, and powerful systems become widely distributed.

Current rules and procedures also need review. The basic rule in subject cataloging is to summarize, not enumerate, the contents of the book. Some feel catalogers spend more time honing down a list of 5 or 6 subject headings than it would take to just include all 5 or 6 subjects in the record. Conceivably, there could be two levels of subject headings--major and minor--or a field could be created in the MARC record in which to list any terms that might be useful. Changes such as these would require changes to the MARC format; more important, however, are the limitations to what LC staff are able to accomplish. Consequently any proposal that could slow down work at LC will require serious thought. One question that needs to be addressed is this: "If LC had more money, would it be better to put it into non-intellectual enhancements (e.g., adding table of contents data) or into subject headings?"

More research on several aspects of this issue is needed.

* * * * * * *

Library of Congress Cataloging in Publication Data
Main entry under title:

Subject access.

"December 1982."
1. Catalogs, On-line--Subject access--Congresses.
2. On-line bibliographic searching--Congresses.
I. Russell, Keith W., 1945- . II. Council on Library Resources.
Z695.S885 1982 025.4'7 82-23563